

10/07/09

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2003, 09:14:16 ; Search time 6001 Seconds
(without alignments)
11634.333 Million cell updates/sec

Title: AJ012098
Perfect score: 2399
Sequence: 1 ATCTTACATGACACACACAA.....TGTAGCAGCAGCTTCTGAG 2399

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_mus:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_ses:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pal:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_ses:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_in:*
- 32: em_htg_mus:*
- 33: em_htg_other:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vtl:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2399	100.0	2399	8	CRE012098	AJ012098 Chlamydom
2	2218	92.5	2522	8	AF289201	AF289201 Chlamydom
3	2119.4	88.3	2454	8	AY055755	AY055755 Chlamydom
4	794.4	33.1	2560	8	AY055756	AY055756 Chlamydom
5	448.6	18.7	2420	8	CF0298228	AJ298228 Chlorella
6	424.6	17.7	1607	8	AF276706	AF276706 Scenedes
7	286.2	11.9	5251	8	DFU07229	U07229 Desulfobibr
8	244.6	10.2	5001	8	SOB271546	AJ271546 Scenedes
9	242.4	10.1	2887	1	DFHYDABG	Y1759 D. fructosov
10	222	9.3	1864	1	DVHYDPE	M27212 D. vulgaris
11	201.4	8.4	2067	1	DVHYDG	X02416 Desulfobibr
12	186	7.8	3265	8	CFU298227	AJ298227 Chlorella
13	180.8	7.5	2585	1	AF120457	AF120457 Megasthae
14	166.6	6.9	15158	1	AE001705	AE001705 Thermotog
15	160	6.7	1940	1	DVHYDC	X57838 D. vulgaris
16	159.4	6.6	1800	1	AF331719	AF331719 Desulfovi
17	148.2	6.2	11577	1	AE013056	AE013056 Thermotana
18	144.8	6.0	1848	3	AF262401	AF262401 Trichomon
19	134	5.6	16133	1	EAC312124	AJ312124 Eubacteri
20	124.4	5.2	2238	1	AF148212	AF148212 Clostridi
21	120.2	5.0	1109	3	AF446077	AF446077 Trichomon
22	109.6	4.6	1109	3	AY028641	AY028641 Trichomon
23	100.8	4.2	1659	3	TVU26964	U26964 Trichomonas
24	89.6	3.7	3625	3	NVHYDROG	Y16775 Nyctotherus
25	86.4	3.6	1544	3	TVU19897	U19897 Trichomonas
26	81.2	3.4	357	3	NOV420000	AJ420000 Nyctother
27	80.2	3.3	113193	1	AF357202	AF357202 Streptomy
28	79.2	3.3	45313	1	SCD95A	AL35743 Streptomy
29	78.2	3.3	34893	1	SC17	AL086743 Streptomy
30	78.2	3.2	125020	9	AF429315	AF429315 Homo sapi
31	77.8	3.2	125020	9	AF429315	AF429315 Homo sapi
32	77.6	3.2	1761	8	AF446076	AF446076 Pityomces
33	76.8	3.2	3113	8	SS95BASAL	AJ001438 Spermatoz
34	76.6	3.2	34611	1	SCD63A	AL356832 Streptomy
35	75.8	3.2	33810	1	SCB6	AL353832 Streptomy
36	75.4	3.1	2467	8	AY033895	AY033895 Neocallim
37	75	3.1	109519	6	AX195929	AX195929 Sequence
38	74	3.1	1389	1	SLI441125	AJ441125 Streptomy
39	74	3.1	1728	6	AX196054	AX196054 Sequence
40	74	3.1	42210	1	SC1C2	AL031124 Streptomy
41	73.6	3.1	30253	1	SC8F4	AL137242 Streptomy
42	73.4	3.1	1109	3	AY028640	AY028640 Trichomon
43	73.4	3.1	12097	1	AE005116	AE005116 Halobacte
44	73	3.0	22991	1	SCD19	AL392149 Streptomy
45	71.8	3.0	25362	1	SCF56	AL133424 Streptomy

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
1	CRE012098	Chlamydomonas reinhardtii mRNA for Fe-hydrogenase.	AJ012098	GI:16945125	Fe-hydrogenase; hyda gene.	Chlamydomonas reinhardtii	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.	1	Kaminski, A.U. and Happe, T.	Isolation and characterization of the hyda gene encoding the Fe-hydrogenase of Chlamydomonas reinhardtii

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2399)
AUTHORS Happe, T.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1998) Happe T., Molecular Biochemistry, Botanical
Institute, Kirschallee 1, D53115-Bonn, GERMANY
REMARK Revised by author 31-JAN-2001
COMMENT Related sequence AJ308413.
FEATURES
Location/Qualifiers
1..2399
/organism="Chlamydomonas reinhardtii"
/db_xref="taxon:3055"
141..1652
/gene="hyda"
/gene="hyda"
141..308
/gene="hyda"
/EC_number="1.18.99.1"
/function="hydrogen evolution"
/codon_start=1
/evidence="experimental"
/product="Fe-hydrogenase"
/protein_id="CAC80065.1"
/db_xref="GI:16945126"
/translation="MSALVTKPCAAYSIRSSCRAROVAPRAPASTVRVALTLEA
PARRLVNACAAAPAPABAPLHVQALAEAPKDDPRKHVCVAVAPRVALVET
LGAPGATTKQALAEGLRLGDEVEPDTLEGADLTMEESSELHLRLHLEHAPSD
EPLPMFTSCPGWIMLEKSPDLIPVSSCKSPOMMLAMVKSYLEKGIAPKDMV
BLSIMPCTRKQSEADDMFCVADPTLRQLDHYITVEIGNIKKEGINLAELEGEM
DNPMVGSGAGVLEFGTGVMAELARVLPFGTPLPRLSISEVNGINLAEKNTNML
VPAKSGFEELHRAAARAPAAHGTPTGRLANDGAGAGTSEBGRGILLRVAVNTL
GNKKLITTKQABEAKIDPEVIMACPAAGCVGGSGPRSDKAITQRQALVNLDEKS
TLRSHENPISRELVDYIGEPGLHKAHELTHYVAGVEKDEKK"
309..1640
/gene="hyda"
/product="Fe-hydrogenase"
BASE COUNT 469 a 745 c 765 g 420 t
ORIGIN
Query Match 100.0%; Score 2399; DB 8; Length 2399;
Best Local Similarity 100.0%; Pred. No. 2,8e-287;
Matches 2399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCTTACATGACACCAAACTCTCGCAGGACCTAGCTCAAACTCTGAAACCTTTT 60
DB 1 ATCTTACATGACACCAAACTCTCGCAGGACCTAGCTCAAACTCTGAAACCTTTT 60
QY 61 TCCACAGTTTACACCCCAATTGCGAGCGCGCTCAAGCTCGCTCCGTTGCTCTTATC 120
DB 61 TCCACAGTTTACACCCCAATTGCGAGCGCGCTCAAGCTCGCTCCGTTGCTCTTATC 120
QY 121 GCACCACTTATTTTCTAATATCGTAGACGCGCAAGATGTGCGGCTCGTGTGAAGC 180
DB 121 GCACCACTTATTTTCTAATATCGTAGACGCGCAAGATGTGCGGCTCGTGTGAAGC 180
QY 181 CCTGCGCGCGCTGTCTATTCGCGGAGCTCTGCGAGGGCGCGGAGCTGCCCCCGGG 240
DB 181 CCTGCGCGCGCTGTCTATTCGCGGAGCTCTGCGAGGGCGCGGAGCTGCCCCCGGG 240
QY 241 CTCGCGTCCGAGCAGACCGTGTAGCGCTTTCGCAACACTTGAAGGCGCGCAGCGC 300
DB 241 CTCGCGTCCGAGCAGACCGTGTAGCGCTTTCGCAACACTTGAAGGCGCGCAGCGC 300
QY 301 GCTTAGGCAACGTCGCTTGGCGCGCTGCGGACCCGCTCGAGGCGCTTTGAGTCATG 360
DB 301 GCTTAGGCAACGTCGCTTGGCGCGCTGCGGACCCGCTCGAGGCGCTTTGAGTCATG 360
QY 361 TCCAGAGGCGCTCGCGAGCTTGCAGAGCCCAAGAGAGACCCCAAGCGCAAGCACTGT 420
DB 361 TCCAGAGGCGCTCGCGAGCTTGCAGAGCCCAAGAGAGACCCCAAGCGCAAGCACTGT 420
QY 421 GCGTGAGGTGCTCGGCGCTTGTGTGTGCTATTGCGCAGACCTTGAGGCTGAGCGCGG 480

DB 421 GCGTGAGGTGCTCGGCGCTTGTGTGTGCTATTGCGCAGACCTTGAGGCTGAGCGCGG 480
QY 481 GCGCCACACCCCCAGAGAGCTGCGGAGAGGCTCCGCGGCTTGAGAGAGTGT 540
DB 481 GCGCCACACCCCCAGAGAGCTGCGGAGAGGCTCCGCGGCTTGAGAGAGTGT 540
QY 541 TTGACACGCTGTTTGGCGCGCAGCTTGACCATATGAGAGAGGAGCAGAGCTCTGACC 600
DB 541 TTGACACGCTGTTTGGCGCGCAGCTTGACCATATGAGAGAGGAGCAGAGCTCTGACC 600
QY 601 GCTTCCCGAGCACTCTGAGAGCCCAACCCGCACTCCGACGAGCGCTGCCATGTTACCA 660
DB 601 GCTTCCCGAGCACTCTGAGAGCCCAACCCGCACTCCGACGAGCGCTGCCATGTTACCA 660
QY 661 GCTGCTGCCCGGCTGGATCGCTATGCTGGAGAAATCTTAACCCGAGCTGATCCCTAGG 720
DB 661 GCTGCTGCCCGGCTGGATCGCTATGCTGGAGAAATCTTAACCCGAGCTGATCCCTAGG 720
QY 721 TGAGCAGCTGCAAGAGCCCCAGATGATGCTGGCGGCATGCTCAAGTCTACTAGCGG 780
DB 721 TGAGCAGCTGCAAGAGCCCCAGATGATGCTGGCGGCATGCTCAAGTCTACTAGCGG 780
QY 781 AAAAGAGGGCATTCGCGCCAAAGACATGCTATGCTGTCATATGCCCTTCACGCGCA 840
DB 781 AAAAGAGGGCATTCGCGCCAAAGACATGCTATGCTGTCATATGCCCTTCACGCGCA 840
QY 841 ACCAGTCGAGGCTACCGCGACTGCTTGTGTGAGAGCGGACCCGACCCCTGCGCGCAGC 900
DB 841 ACCAGTCGAGGCTACCGCGACTGCTTGTGTGAGAGCGGACCCGACCCCTGCGCGCAGC 900
QY 901 TGACCAAGCTCATCAACACCGTGAAGCTGGGCAACATCTTCAAGAGCGCGCATCAAC 960
DB 901 TGACCAAGCTCATCAACACCGTGAAGCTGGGCAACATCTTCAAGAGCGCGCATCAAC 960
QY 961 TGGCGGAGCTGCCCGAGGGCGAGTGGACAAATCCAAATGGCGCTGGCGCGCGCGG 1020
DB 961 TGGCGGAGCTGCCCGAGGGCGAGTGGACAAATCCAAATGGCGCTGGCGCGCGCGG 1020
QY 1021 TGCCTGTTGGGACCAACCGGCGGTGCATGAGAGCGCGCTGCGCAGGCTTATGAGCTGT 1080
DB 1021 TGCCTGTTGGGACCAACCGGCGGTGCATGAGAGCGCGCTGCGCAGGCTTATGAGCTGT 1080
QY 1081 TCACGGGCAAGCGCGCTGCCGCGCTGAGCTGAGCGAGGTGGCGCATGAGCGGCATCA 1140
DB 1081 TCACGGGCAAGCGCGCTGCCGCGCTGAGCTGAGCGAGGTGGCGCATGAGCGGCATCA 1140
QY 1141 AGGAGACCAATCACTACATGGTGCCGCGCGCGCTCCAGTTTGAAGAGCTGTGAAGC 1200
DB 1141 AGGAGACCAATCACTACATGGTGCCGCGCGCGCTCCAGTTTGAAGAGCTGTGAAGC 1200
QY 1201 ACCGCGCGCGCGCGCGCGCGAGAGCGCGCGGCAAGGACCCCGGCGCTGAGCTGGG 1260
DB 1201 ACCGCGCGCGCGCGCGCGCGAGAGCGCGCGGCAAGGACCCCGGCGCTGAGCTGGG 1260
QY 1261 AGGCGGCGCGGCGCTTCAACAGCGAGAGCGGAGGCGGCGCATCACTGCGCGTGGCGG 1320
DB 1261 AGGCGGCGCGGCGCTTCAACAGCGAGAGCGGAGGCGGCGCATCACTGCGCGTGGCGG 1320
QY 1321 TGGCCAAAGGCGTGGGCAAGCGCAAGAGCTGATCAACCAATGAGAGCGCGGCGAGCA 1380
DB 1321 TGGCCAAAGGCGTGGGCAAGCGCAAGAGCTGATCAACCAATGAGAGCGCGGCGAGCA 1380
QY 1381 AGTACGACTTTGTGAGATCATGAGCTTGCCTCGCGGCGCTGTGTGAGCGCGCGCGCAGC 1440
DB 1381 AGTACGACTTTGTGAGATCATGAGCTTGCCTCGCGGCGCTGTGTGAGCGCGCGCGCAGC 1440
QY 1441 CCGGCTTCCACCGACAAAGGCTATCAGCGAAGCGCGGCGGCGCTGTACAACTTGAGCG 1500
DB 1441 CCGGCTTCCACCGACAAAGGCTATCAGCGAAGCGCGGCGGCGCTGTACAACTTGAGCG 1500
QY 1501 AGAAGTCCAGCGCTGCGCGAGCGCAGCAAGACCGTGCATCGGAGCTGTACGACACGT 1560

Db 1501 AGAATCCACGCTGCCCCGAGCCACGAGAACCCGTCATCCCGAGCTGTACACACGT 1560
Qy 1561 ACCTCGGAGAGCGCTGCGCCACAGAGCGCAGAGCTGTGACACCCACTAGCTGCGCC 1620
Db 1561 ACCTCGGAGAGCGCTGCGCCACAGAGCGCAGAGCTGTGACACCCACTAGCTGCGCC 1620
Qy 1621 GCGGCGTGAAG 1680
Db 1621 GCGGCGTGAAG 1680
Qy 1681 CAGCTTCAAG 1740
Db 1681 CAGCTTCAAG 1740
Qy 1741 GGTGCTCTCTTAT 1800
Db 1741 GGTGCTCTCTTAT 1800
Qy 1801 GAGTGTGTACACAGCATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1801 GAGTGTGTACACAGCATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1861 GATTGAGGCTAGTACTGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 GATTGAGGCTAGTACTGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy 1921 CCAGGACAGCTGCGAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 CCAGGACAGCTGCGAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1981 GATATATTGAAGCTGACAAACCTGCACTTATTTGCTTTTACAGTGGCCCAATCTTG 2040
Db 1981 GATATATTGAAGCTGACAAACCTGCACTTATTTGCTTTTACAGTGGCCCAATCTTG 2040
Qy 2041 GTTGAAGCTTAAACATGTTTGGGAAACATTTCACTTAAACGTTGGGGGCTTGAAGA 2100
Db 2041 GTTGAAGCTTAAACATGTTTGGGAAACATTTCACTTAAACGTTGGGGGCTTGAAGA 2100
Qy 2101 TGGGACGTTGTGCGCTGAGTGGGCGGGGAACTGGGAGATTTAGGCTAGTGGCA 2160
Db 2101 TGGGACGTTGTGCGCTGAGTGGGCGGGGAACTGGGAGATTTAGGCTAGTGGCA 2160
Qy 2161 TACGACAAAGGAGCGGCGCTGAGATTTGAGCACTTACGCGGAACTTATGAACGTAGCGCT 2220
Db 2161 TACGACAAAGGAGCGGCGCTGAGATTTGAGCACTTACGCGGAACTTATGAACGTAGCGCT 2220
Qy 2221 TTTATACCAACCTGATGAGTGAAGCTTTGTTGAGGACCAAGCGGTTAGGAAGCGGAGA 2280
Db 2221 TTTATACCAACCTGATGAGTGAAGCTTTGTTGAGGACCAAGCGGTTAGGAAGCGGAGA 2280
Qy 2281 GATGCAATTGCAACCGCTGTAAAGAAAGCGCATAGCTAGACACTGTATGTGAGACC 2340
Db 2281 GATGCAATTGCAACCGCTGTAAAGAAAGCGCATAGCTAGACACTGTATGTGAGACC 2340
Qy 2341 TTGGCGCAGCAG 2399
Db 2341 TTGGCGCAGCAG 2399

RESULT 2
AF289201 2522 bp mRNA linear PLN 17-AUG-2000
LOCUS Chlamydomonas reinhardtii Fe-hydrogenase precursor (hvd1) mRNA,
DEFINITION complete cds; nuclear gene for chloroplast product.
ACCESSION AF289201.1 GI:9837539
VERSION AF289201.1
KEYWORDS Chlamydomonas reinhardtii.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 2522)
AUTHORS Metz, L.J.

TITLE The iron hydrogenase of Chlamydomonas reinhardtii has a single
JOURNAL folding domain containing an H-cluster catalytic center and lacking
REFERENCES bound electron carriers
AUTHORS Unpublished
TITLE 2 (bases 1 to 2522)
JOURNAL Metz, L.J.
DIRECT SUBMISSION
SUBMITTED (24-JUN-2000) Molecular Genetics and Cell Biology, The
UNIVERSITY OF CHICAGO, 1103 E. 57th St., Chicago, IL 60637, USA
FEATURES
source 1.2522
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/map="LG III"
/clone="CLS8b09"
1.2522
/gene="hvd1"
135.1628
/gene="hvd1"
/EC number="1.18.99.1"
/note="Method: conceptual translation with partial peptide
sequencing"
/codon_start=1
/product="Fe-hydrogenase precursor"
/protein_id="AAG00591.1"
/db_xref="GI:9837540"
/translation="MSALVLRKPAASVIRGSSCPARQVAPAPLAATRVVALTLEA
PARRIGNVACAAAPAEAPLPSHVOALBELKPKDDPRKIVCQVAPARVALIET
LGLAPGATLTKQALAGLRLGFDEVFTLFGADLTIMEGSELRLHRLTSLAPHS
BPLPMTSCPCPMIMLEKSYDPLIPYSSCSQPMMLAAYKSLAEKXGLAPKDMV
AVSIMPTCKROSEADRMFCVADPLRLDHLVITVELGNIFKEKGINLAELPEGE
DMPKSGSAGVLPFTGSMVEMALRTALFTPLPLPLSLSEVRGMLGKETNTM
VAPRSKFEELIKHRAARARAEAACTPEPLAMDGAGCTSDRGGITLRVAVANGL
GNAKULITMQRGAEAKIDPEVETMACPAGVGGGQPRSTDAITQRQALYNLDEKS
TLRSHENPSIRELYDTYLGEPHKAHELTHYVAGVEKEKK"
transit_peptide 135.302
/gene="hvd1"
303.1625
mat_peptide /gene="hvd1"
/product="Fe-hydrogenase"
BASE COUNT 484 a 777 c 812 g 449 t
ORIGIN
Query Match 92.5%; Score 2218; DB 8; Length 2522;
Best Local Similarity 94.2%; Pred. No. 6, 2e-265;
Matches 2375; Conservative 0; Mismatches 0; Indels 147; Gaps 1;
Qy 25 CTGCGAGCACTAGCTCAAAACCTTCTTCCAGAGTTTACACCCCAATTGG 84
Db 1 CTGCGAGGACACTAGCTCAAAACCTTCTTCCAGAGTTTACACCCCAATTGG 60
Qy 85 GACGCGCTCAAGCTGCTGCTGCTCTTCAATGACCAACCTATTAATTCTAATATC 144
Db 61 GACGCGCTCAAGCTGCTGCTGCTCTTCAATGACCAACCTATTAATTCTAATATC 120
Qy 145 GTAGAGCGCAAGAGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204
Db 121 GTAGAGCGCAAGAGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Qy 205 GCAAGCTCTGCAAGGAGGCGGAGAGTGCAGCCCGCGCTGCTGCGAGCAGCAGCGTGC 264
Db 181 GCAAGCTCTGCAAGGAGGCGGAGAGTGCAGCCCGCGCTGCTGCGAGCAGCAGCGTGC 240
Qy 265 GTGTAGCCCTTGCAACACTTTGAGGCGCCGCAAGCCGCTTAGGCAACGTTGCTGGCGG 324
Db 241 GTGTAGCCCTTGCAACACTTTGAGGCGCCGCAAGCCGCTTAGGCAACGTTGCTGGCGG 300
Qy 325 CTGCGGCAACCCCTGCGGAGGCGGCTTTAGTCATGTCCAGAGAGGCTGCGGAGACTTG 384
Db 301 CTGCGGCAACCCCTGCGGAGGCGGCTTTAGTCATGTCCAGAGAGGCTGCGGAGACTTG 360
Qy 385 CCAAGCCCAAGAGCAAGCCAGCGGCAAGCAAGTGTGAGAGTGTGCTGCGGCGCTTC 444

```
Db 361 CCAAGCCCAAGAGACACCCCAACGGCAAGACGTCGTGTCAGATGGCTCCGCGCTTC 420
Qy 445 GTGTGCTATTGCGAGAGACCTTGAGCCCTGAGCGCGAGGCCACCAACCCCAAGAGGCTGG 504
Db 421 GTGTGCTATTGCGAGAGACCTTGAGCCCTGAGCGCGAGGCCACCAACCCCAAGAGGCTGG 480
Qy 505 CCGAGAGGCTTCGCGCGCTCGGCTTTTGAAGAGTGTTTTGAACAGCTGTTTGGCCCGCAGC 564
Db 481 CCGAGAGGCTTCGCGCGCTCGGCTTTTGAAGAGTGTTTTGAACAGCTGTTTGGCCCGCAGC 540
Qy 565 TGACCATCATGAGAGAGAGGCGAGAGCTGTGACACCCGCTTACACAGAGCTGAGCTGATCGCTA 624
Db 541 TGACCATCATGAGAGAGAGGCGAGAGCTGTGACACCCGCTTACACAGAGCTGAGAGGCC 600
Qy 625 ACCGCACTCCGAGAGAGCGCTGCGCATGTTCACACAGCTGTGCGCGGCTGATCGCTA 684
Db 601 ACCGCACTCCGAGAGAGCGCTGCGCATGTTCACACAGCTGTGCGCGGCTGATCGCTA 660
Qy 685 TGCTGAGAGAAATCTTACCCGAGACTGATCCCTTACGTGAGAGAGCTGCAAGAGCCCGAGA 744
Db 661 TGCTGAGAGAAATCTTACCCGAGACTGATCCCTTACGTGAGAGAGCTGCAAGAGCCCGAGA 720
Qy 745 TGATGCTGAGCGCGCATGATGATGCTTACCTAGGAGAAAGAGAGGATGGCGCAAG 804
Db 721 TGATGCTGAGCGCGCATGATGATGCTTACCTAGGAGAAAGAGAGGATGGCGCAAG 780
Qy 805 ACATGTCATGATGTGTTCATCATGCTGCAACGCGCAAGAGCTGAGAGCTGACCGGACT 864
Db 781 ACATGTCATGATGTGTTCATCATGCTGCAACGCGCAAGAGCTGAGAGCTGACCGGACT 840
Qy 865 GGTCTGTGTGAGAGAGCGGACCCCAACCTGCGCGACAGCTGAGACACGTATATCAACGCTGG 924
Db 841 GGTCTGTGTGAGAGAGCGGACCCCAACCTGCGCGACAGCTGAGACACGTATATCAACGCTGG 900
Qy 925 AGCTGGGCAACATCTTCAAGAGAGCGGAGCATCAACCTGCGAGCTGCGGAGGGGAGT 984
Db 901 AGCTGGGCAACATCTTCAAGAGAGCGGAGCATCAACCTGCGAGCTGCGGAGGGGAGT 960
Qy 985 GGGCAATTCGAATGAGGCGTGGGCTCGGAGCGCGCGCTGTCTTTCGAGCACCAGCGCGTG 1044
Db 961 GGGCAATTCGAATGAGGCGTGGGCTCGGAGCGCGCGCTGTCTTTCGAGCACCAGCGCGTG 1020
Qy 1045 TCATGAGAGGCGCGCTCGGACCGGCTTATGAGCTGTTCACGCGGACGCGCTGCGCGCT 1104
Db 1021 TCATGAGAGGCGCGCTCGGACCGGCTTATGAGCTGTTCACGCGGACGCGCTGCGCGCT 1080
Qy 1105 TGAAGCTGAGAGAGTSCGCGGATGAGAGGAGCATCAAGAGACCAACATCAACATGCTG 1164
Db 1081 TGAAGCTGAGAGAGTSCGCGGATGAGAGGAGCATCAAGAGAGACCAACATCAACATGCTG 1140
Qy 1165 CCGGCGCGCGGCTCCAGTTTGAAGAGCTGCTGAAGCAACGCGCGCGCGCGCGCGAG 1224
Db 1141 CCGGCGCGCGGCTCCAGTTTGAAGAGCTGCTGAAGCAACGCGCGCGCGCGCGCGAG 1200
Qy 1225 CCGGCGCGGACCGGACCCCGGCGCTGCTGAGAGCGGCGGCGGCTTACCAAGG 1284
Db 1201 CCGGCGCGGACCGGACCCCGGCGCTGCTGAGAGCGGCGGCGGCTTACCAAGG 1260
Qy 1285 AGAAGCGAGAGGCGGAGTCACTGCGCTGCGCTGCGCAACGCGCTGAGGAGCGCA 1344
Db 1261 AGAAGCGAGAGGCGGAGTCACTGCGCTGCGCTGCGCAACGCGCTGAGGAGCGCA 1320
Qy 1345 AGAAGCTGATCACCAGATGCAAGGCGGAGCGGAGCGCAAGTTCGACTTTGTGAGATCATG 1404
Db 1321 AGAAGCTGATCACCAGATGCAAGGCGGAGCGGAGCGCAAGTTCGACTTTGTGAGATCATG 1380
Qy 1405 CTTGCGCGCGGAGCTGTGTGAGGCGGCGGAGCGGCGGCTTACCAAGGAGCATG 1464
Db 1381 CTTGCGCGCGGAGCTGTGTGAGGCGGCGGAGCGGCGGCTTACCAAGGAGCATG 1440
Qy 1465 CCGAGAAAGCGGAGGCGGCTGTACAACTTGAGAGAGAGTCAACGCTGCGCGGAGCC 1524
Db 1441 CCGAGAAAGCGGAGGCGGCTGTACAACTTGAGAGAGAGTCAACGCTGCGCGGAGCC 1500
Qy 1525 ACAGAAACCGGTCATCCGAGAGCTGTACAGACAGTAACTCGGAGAGCCGCTGGGCGACA 1584
Db 1501 ACAGAAACCGGTCATCCGAGAGCTGTACAGACAGTAACTCGGAGAGCCGCTGGGCGACA 1560
Qy 1585 AGGCGACAGAGCTGCTGACACCCACTACGATGAGCGCGGCGGCTGAGAGAGAGAGAGA 1644
Db 1561 AGGCGACAGAGCTGCTGACACCCACTACGATGAGCGCGGCGGCTGAGAGAGAGAGAGA 1620
Qy 1645 AGAAGTGAAGAGGCGGAGAGGCTCTTTGGCGGAGACAGCTTCAAGAGAGAGGCGAT 1704
Db 1621 AGAAGTGAAGAGGCGGAGAGGCTCTTTGGCGGAGACAGCTTCAAGAGAGAGGCGAT 1680
Qy 1705 TAGCACTACCGTAATATGACATGATGAGGAGTTCGCGGAGTCTCTTATATGAAATG 1764
Db 1681 TAGCACTACCGTAATATGACATGATGAGGAGTTCGCGGAGTCTCTTATATGAAATG 1740
Qy 1765 GGTCAAAATAGGCGGAGGTCAAATGTTTCTTTTGAAGTGTGTACAGATGAGGCGAC 1824
Db 1741 GGTCAAAATAGGCGGAGGTCAAATGTTTCTTTTGAAGTGTGTACAGATGAGGCGAC 1800
Qy 1825 GTGTGCGAGCGCAG----- 1839
Db 1801 GTGTGCGAGCGCAGTTGCTTCAGTGCACGCGCTCCGCTGTGTGCGCAGCTGCGCT 1860
Qy 1840 ----- 1839
Db 1861 TGATATATCACCGGAGAGATTTATGAAAGGGGAGCTCAAGAAAGCTCATATTGAGC 1920
Qy 1840 -----TAGGCTGTTCATGACG 1857
Db 1921 AATGCTGTCTCTTCCACATGATGTGTGAGCGGCTCCGCAATGAGCTGTTCATGACG 1980
Qy 1858 CTGGCAATTAGGGTATGATCTGSCATGAGGAGCGCGCTTCTTAACGAAATGGGCTATC 1917
Db 1981 CTGGCAATTAGGGTATGATCTGSCATGAGGAGCGCGCTTCTTAACGAAATGGGCTATC 2040
Qy 1918 CCTCAGGAGCAGTGGGAATGCGGCTGCCATCAACGCAAAATTTCTGGCTTCAATGCT 1977
Db 2041 CCTCAGGAGCAGTGGGAATGCGGCTGCCATCAACGCAAAATTTCTGGCTTCAATGCT 2100
Qy 1978 TCTGATATTTGAAGCTGCAAAACCTGCATTTCTATTGCTGTTTACAGAGTCCCAATC 2037
Db 2101 TCTGATATTTGAAGCTGCAAAACCTGCATTTCTATTGCTGTTTACAGAGTCCCAATC 2160
Qy 2038 TTGGTTGAAGCTAAACATGTTGGGAACAATCTTAACTTAAGCGGTGGGGTTGA 2097
Db 2161 TTGGTTGAAGCTAAACATGTTGGGAACAATCTTAACTTAAGCGGTGGGGTTGA 2220
Qy 2098 GGATGCGCAGCTTGTGCGCTGTGTGGGCGGAGAGCTGGGTAGCATTTAGGCTAGCTG 2157
Db 2221 GGATGCGCAGCTTGTGCGCTGTGTGGGCGGAGAGCTGGGTAGCATTTAGGCTAGCTG 2280
Qy 2158 GCATACGACAAAGGGGCGCGGAGATGAGACTTGAAGCTTGAAGCTTGAAGAGTAC 2217
Db 2281 GCATACGACAAAGGGGCGCGGAGATGAGACTTGAAGCTTGAAGCTTGAAGAGTAC 2340
Qy 2218 GCTTATATCCCACTGATGCAATGACGTTGAGGAGCAACAGGCGGTAGGAAGGCGG 2277
Db 2341 GCTTATATCCCACTGATGCAATGACGTTGAGGAGCAACAGGCGGTAGGAAGGCGG 2400
Qy 2278 AGAGATGATTCGAAACCCCTGTAAAGACGAGTACGTAAGTATGAGTGA 2337
Db 2401 AGAGATGATTCGAAACCCCTGTAAAGACGAGTACGTAAGTATGAGTGA 2460
Qy 2338 CCTTGGGCGAGCGCAGAGAGAGGAGTGCATACGCGCTTGAAGCAGAGCACTTCTG 2397
Db 2461 CCTTGGGCGAGCGCAGAGAGAGGAGTGCATACGCGCTTGAAGCAGAGCACTTCTG 2520
Qy 2398 AG 2399
Db 2521 AG 2522
```

LOCUS	AY055755	2454 bp	mRNA	linear	PLN 31-DEC-2001
DEFINITION	Chlamydomonas reinhardtii iron-hydrogenase Hyda (hyda) mRNA.				
ACCESSION	AY055755				
VERSION	AY055755				
KEYWORDS	AY055755.1 GI:18026269				
SOURCE	Chlamydomonas reinhardtii.				
ORGANISM	Chlamydomonas reinhardtii				
REFERENCE	Chlamydomonas reinhardtii				
AUTHORS	Eukaryota, Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.				
REFERENCE	1 (bases 1 to 2454)				
AUTHORS	Forestier, M., Zhang, L., Plummer, S., Ahmann, D., Seibert, M. and Ghirardi, M.L.				
TITLE	Two putative Fe-only hydrogenases cloned from Chlamydomonas reinhardtii are coexpressed in cells undergoing anaerobiosis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2454)				
AUTHORS	Forestier, M., Zhang, L., Plummer, S., Ahmann, D., Seibert, M. and Ghirardi, M.L.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-SEP-2001) Basic Sciences, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401, USA				
FEATURES	Location/Qualifiers				
source	1..2454				
gene	/organism="Chlamydomonas reinhardtii"				
CDS	/strain="21gr"				
	/db_xref="taxon:3055"				
	/clone="lam9-1"				
	1..2454				
	/gene="hyda"				
	120..1613				
	/gene="hyda"				
	/codon_start=1				
	/product="iron-hydrogenase Hyda"				
	/protein_id="PAL23572.1"				
	/db_xref="gi:18026270"				
	/translation="MSALVLPKCAVSIIRGSSCARQVAPRAPAASTVRVALATLEA PARRGNVCAAAAPAAEPAPSPHVOALAEIAKPDPTKHCVCQVAPAVRIATETD LGILGATPTKQIAEGLRRIGDEVEFDLTPADLTIMESSEGLHRLTLEHLEAHPHSD EPLPMFTSCCPGSIAMLEKSTPDLIPYVSSCKSQOMLAAVYKSLAEKGIARXDNV MWSIMPTRKQSEADRWFCVDADPTLRALDHTVTVYELGNI FFERGINIALEPESGW DNVSGGAGVILFGTTGGVMEALRTYVELTPPLPRLSLSEVRGNDGDKETNITMT VPAPSKFEELIKHRAARAPAAAHGTPGLAMPDGAGFTSEDRGGITLRVAVANGI GNARKLITKMOAGAKYDFVFIACPGACGCGGGGOPSIDKAIITOKQOALVYNDEKS TLRSHENPSTRELVDYTLGEPFLGHKAEHLHTHYVAGVERKEEK"				
BASE COUNT	493 a 747 c 784 g 430 t				
ORIGIN					
Query Match	88.3%; Score 2119.4; DB 8; Length 2454;				
Best Local Similarity	93.7%; Pred. No. 9.2e-253;				
Matches 2280; Conservative	0; Mismatches 6; Indels 147; Gaps 1;				
40	CTCAAACCTCGAAACCTTTTTCACAAGTTATTCACCCCAATTGCGAGCCGCTCGAACG	99			
1	CTCAAAACCTCGAAACCTTTTTCACAAGTTATTCACCCCAATTGCGAGCCGCTCGAACG	60			
100	TGCTCCGTTGCTCTTTCATCGCACACCACTATTATTCTAATATCTGTAGAGCGGACAGA	159			
61	TGCTCCGTTGCTCTTTCATCGCACACCACTATTATTCTAATATCTGTAGAGCGGACAGA	120			
160	TGTGGGCGCTGCTGTCTGAAGCCCTGCGGGCGGCGTCTATTTCGCGGAGCTCTGACAGG	219			
121	TGTGGGCGCTGCTGTCTGAAGCCCTGCGGGCGGCGTCTATTTCGCGGAGCTCTGACAGG	180			
220	CGCGGAGGTGCGCCCGCGGCTCCGCTTCGACGCCAGCAGCCGTGCGTATGCAACCTTGC	279			
181	CGCGGAGGTGCGCCCGCGGCTCCGCTTCGACGCCAGCAGCCGTGCGTATGCAACCTTGC	240			
280	CACTTGAGCGCGCGGACGCGGCTTAGGACAGCTGCTTGGCGGCGGCTCCGACACCGGCTG	339			

Db	241	CAC TTGAGGGGCCCGCAGCCGCTTAGGCAACGTCCTTGCGCGGCTGCGCAACCCGCTG	300
Oy	340	CGAGAGGGGCGCTTTTAAATCATATGTCAGAGAGGCGCTGCGGAGCTTTGCCAACCCCAAGAG	399
Db	301	CGAGGGGCGCTTTTAAATCATATGTCAGAGAGGCGCTGCGGAGCTTTGCCAACCCCAAGAG	360
Oy	400	ACCCACGCGCAACGATCTGACGTGAGAGGTGCGCTCGGAGCGTTGCGTGTGCTATTGCG	459
Db	361	ACCCACGCGCAACGATCTGACGTGAGAGGTGCGCTCGGAGCGTTGCGTGTGCTATTGCG	420
Oy	460	AGACCTTGAGGCTTGCGCCCGGCGCCCAACCCCAAGACAGTGGCCGAGGCGCTCCGCC	519
Db	421	AGACCTTGAGGCTTGCGCCCGGCGCCCAACCCCAAGACAGTGGCCGAGGCGCTCCGCC	480
Oy	520	GCGTGGCTTTGACGAGGTGTTTGACACGCTGTTTGGCGCCGACCTGACATCATGAGG	579
Db	481	GCGTGGCTTTGACGAGGTGTTTGACACGCTGTTTGGCGCCGACCTGACATCATGAGG	540
Oy	580	AGGGACACGAGCTGTGTCCACCGGCTTCACCGACACCTGGAGGGCCACCCGACCTCGAG	639
Db	541	AGGGACACGAGCTGTGTCCACCGGCTTCACCGACACCTGGAGGGCCACCCGACCTCGAG	600
Oy	640	AGCGGCTGCCATTTTACAGAGCTGCTGCGCCCGGCTGGATCGCTATCTGGAGAAATCTT	699
Db	601	AGCGGCTGCCATTTTACAGAGCTGCTGCGCCCGGCTGGATCGCTATCTGGAGAAATCTT	660
Oy	700	ACC GGACCTGATCCCTTACGTGAGCTGACAGCTGCAAGAGACCCCCAGATATGCTGGCGGCA	759
Db	661	ACC GGACCTGATCCCTTACGTGAGCTGACAGAGACCCCCAGATATGCTGGCGGCA	720
Oy	760	TGGTCAGTCTCTACTTAGCGGGAAGAGAGGCAATCGGCGCAAGAGATGTCATGCTGT	819
Db	721	TGGTCAGTCTCTACTTAGCGGGAAGAGAGGCAATCGGCGCAAGAGATGTCATGCTGT	780
Oy	820	CCATCATATGCCCTTGACACCGCAAGAGAGTTCGAGAGCTGACCCGACCTGCTGTGAGC	879
Db	781	CCATCATATGCCCTTGACACCGCAAGAGAGTTCGAGAGCTGACCCGACCTGCTGTGAGC	840
Oy	880	CCGACCCCAACCTTGCGCACAGCTGAGCCACGTCATCACACCGTGGAGCTGGGCAACATCT	939
Db	841	CCGACCCCAACCTTGCGCACAGTGTGAGCCACGTCATCACACCGTGGAGCTGGGCAACATCT	900
Oy	940	TCAAGAGAGCGGTGATCAACTTGGCCGAGCTGCCCGAGGGCGAGTGGGCAATTCATG	999
Db	901	TCAAGAGAGCGGTGATCAACTTGGCCGAGCTGCCCGAGGGCGAGTGGGCAATTCATG	960
Oy	1000	GCGTGGGCTGGGGGCGCGCGTGCTGTTCGACACACCGCGGTGATGAGAGCGCGCG	1059
Db	961	GCGTGGGCTGGGGGCGCGCGTGCTGTTCGACACACCGCGGTGATGAGAGCGCGCG	1020
Oy	1060	TGCGCACGGGCTTATGAGCTGTTTCAACGGGCAACGCGCTGCGCGGCTGAGCTGACGAGG	1119
Db	1021	TGCGCACGGGCTTATGAGCTGTTTCAACGGGCAACGCGCTGCGCGGCTGAGCTGACGAGG	1080
Oy	1120	TGCGCGGATGAGAGCGGATCAAGAGAGCAACATACATATGATGCTCCGCGCGGGTCCA	1179
Db	1081	TGCGCGGATGAGAGCGGATCAAGAGAGCAACATACATATGATGCTCCGCGCGGGTCCA	1140
Oy	1180	AGTTTGAAGAGCTGTGAAGACACCGCGCGCGCGCGCGCGCGCGAGCGCGCGGACCGCA	1239
Db	1141	AGTTTGAAGAGCTGTGAAGACACCGCGCGCGCGCGCGCGCGCGAGCGCGCGGACCGCA	1200
Oy	1240	CCCCCGGGCGCTTGAGAGCGGCGCGCGCGGCTTACACAGCGAGAGACGGAGCGGCG	1299
Db	1201	CCCCCGGGCGCTTGAGAGCGGCGCGCGCGGCTTACACAGCGAGAGACGGAGCGGCG	1260
Oy	1300	GCATCACTGCGCGTGGCGGTGGCCAAACGGGCTGGGCAACGCCAAGAGCTGATCAACA	1359
Db	1261	GCATCACTGCGCGTGGCGGTGGCCAAACGGGCTGGGCAACGCCAAGAGCTGATCAACA	1320
Oy	1360	AGATGACAGCGCGGAGAGCCAAAGTACATTTTGGAGATCATATGGCTTGCCCGCGGCT	1419

[illegible]

DB	1633	AGGCGGCGCCGAGG	1646
RESULT 5	CFU298228	2420 bp	mRNA linear
LOCUS	CFU298228		PLN 04-JAN-2002
DEFINITION	Chlorrella fusca mRNA for Fe-hydrogenase (hyla gene).		
ACCESSION	AJ298228		
VERSION	AJ298228.1		
KEYWORDS	Fe-hydrogenase; hyla gene.		
SOURCE	'Chlorrella' fusca.		
ORGANISM	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales; Scenedesmeceae; Scenedesmus.		
REFERENCE	1 Winkler, M.		
TITLE	The Fe-hydrogenase of chlorrella fusca		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2420)		
AUTHORS	Happe, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-DEC-2000) Happe T., Molekulare Biochemie, Botanisches Institut der Universitaet Bonn, Karlrobert-Kreien-Str.13, Bonn, NRW, 53115, GERMANY		
FEATURES	Location/Qualifiers		
source	1..2420		
gene	/organism="Chlorrella' fusca"		
CDS	/db_xref="taxon:3073"		
	67..1377		
	/gene="hyla"		
	67..1377		
	/gene="hyla"		
	/EC_number="1.18.99.1"		
	/function="hydrogen metabolism"		
	/codon_start=1		
	/product="Fe-hydrogenase"		
	/protein_id="CAC83291.1"		
	/db_xref="GI:18073435"		
	/translation="MCCPVVAASHRAGRAHVAARAAGPTSECDPTPOAKLPHMOQA LDELAKESRRLMIAOIASAVRAIAETTGILPAGDTIGOLYGLMIGPDYEDTL FGAULTIMESTELHRLQDLBHPKPEKLPMPGTSCCGVMVMEKSNELPYLS SCSPQMLGAVIKNYIAQOVGPSOICVANSVMPCVKQGEADREWFNTTGAGLARI VDHYTTAEVGGKIFLERGILKLNELPESNPONPIGEIGGALLPCTGGVMEALRTVY EVAVQKMGVDEFEVEVLEGLEIKAEITLKPQDS PKAFAGADGCIITLKIAVANGI GNAKLILKSLSEGGAKYDPIEVNACPGCGCGGQSPSTDKOILQKQQAAMVNLDRS TIRSHNPTPIQALYDKFLGAPRSHKADLLHHTHYVAGIPEEK"		
transit_peptide	67..129		
mat_peptide	/gene="hyla"		
	130..1374		
	/gene="hyla"		
	/product="Fe-hydrogenase"		
	/EC_number="1.18.99.1"		
BASE COUNT	578 a 581 c 665 g 592 t 4 others		
ORIGIN			
Query Match	18.7%; Score 448.6; DB 8; Length 2420;		
Best Local Similarity	61.7%; Pred. No. 1.6e-46;		
Matches	807; Conservative 0; Mismatches 409; Indels 93; Gaps 2;		
QY	331 CACCCGCTGCGAGCGCCTTTGATGATCTCCAGACAGCGCTGCGAGCTTCCCAAGC	390	
DB	158 CTCCAACCTCAGAGCCCAAGCTGCTCAGCGACGAGCGCTTGATGAGCTCCCAAGC	217	
QY	391 CCAAGGACGACCCCAAGCGCAAGCAGCTTCCGCTGACAGGTGCTCCGCGCTTGTGTGCG	450	
DB	218 CCAAGGA-----GAGCAGGAGGTTGATGATCGGCAAAATCGCCTCCGCTGTGTGTGCG	271	
QY	451 CTATTGCGGAGACCTGAGCGCTGCGCGCGCGCCACACATCCCAAGCAGCTGCGCGAGG	510	
DB	272 CTATTGCTGAGACCATTTGCTTGAGCCCAAGAGATGTCACCATTTGGGCACTCGTGAAGT	331	
QY	511 GCCTTCGCGCGCTGCGCTTTGACGAGGTGTTTGACAGCTGTTTGGCGCGCAGCTGACCA	570	


```

Db      332 GGGTGGCTATGCTTGGCTTGAATTATGCTTTGACACCCCTGTTGGTGTGACCTGACCA 391
Qy      571 TCATGAGAGAGGAGGAGGAGCTGTGACCGGCTCAACGAGCACTTGGAGGCCACCGC 630
Db      392 TTATGAGAGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGACCCCA 451
Qy      631 ACTCCAGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 690
Db      452 ACAAGAGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 511
Qy      691 AGAATCTTACCCGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 750
Db      512 AAAAGAGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 571
Qy      751 TGGGAGGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 810
Db      572 TGGGAGGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 631
Qy      811 TCATGAGTGTCCATGATGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGGAG 870
Db      632 GCAACGAGTGTGATGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGGAGGAG 691
Qy      871 GTGTGAGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 930
Db      692 ACAACAGAGTGTGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 751
Qy      931 GCAACATCTTCAAGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAG 990
Db      752 GTAAAGATATCTGTGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAG 811
Qy      991 ATCAATGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 1050
Db      812 ACCCATTTGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 871
Qy      1051 AGGAGGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAGG 1110
Db      872 AGGAGAGCTTGGAGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGG 931
Qy      1111 TGAGAGGAGGAGGAGGAGGAGCTGTGACCGGCTGAGAGCACTTGGAGAGGAG 1170
Db      932 TTGAGAGGAGGAGGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAGGAG 991
Qy      1171 CCGGAGTCAAGTTTGGAGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAG 1230
Db      992 ACCAGAGCCCATTCAGAGCTT----- 1013
Qy      1231 CGCAGGAGGAGGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAGGAG 1290
Db      1014 -----CGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1024
Qy      1291 GCAGGAGGAGGAGGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAGGAGG 1350
Db      1025 ATGAGGAGGAGGAGGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAGGAG 1084
Qy      1351 TGATACCAAGATGAGGAGGAGGAGGAGCTGTGAGAGCACTTGGAGAGGAGGAG 1410
Db      1085 TCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1144
Qy      1411 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1470
Db      1145 CTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1204
Qy      1471 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1530
Db      1205 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1264
Qy      1531 ACCGATCATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1590
Db      1265 ACCCATTCATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1324
Qy      1591 ACAGAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1639
Db      1325 ATGATGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1373

```

```

RESULT 6
AF276706
LOCUS AF276706 1607 bp mRNA linear PLN 15-MAR-2001
DEFINITION Scenedesmus obliquus Fe-hydrogenase mRNA, partial cds.
ACCESSION AF276706
VERSION AF276706.1 GI:12581497
KEYWORDS
SOURCE
ORGANISM Scenedesmus obliquus.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlorococcales; Scenedesmaceae; Scenedesmus.
REFERENCE
1 (bases 1 to 1607)
Munschiers, R., Stangier, K., Senger, H. and Schulz, R.
Molecular evidence for a Fe-hydrogenase in the green alga
Scenedesmus obliquus
JOURNAL Curr. Microbiol. 42 (5), 353-360 (2001)
MEDLINE 21292711
PUBMED 11400057
REFERENCE 2 (bases 1 to 1607)
Munschiers, R., Stangier, K., Senger, H. and Schulz, R.
Direct Submission
AUTHORS Submitted (09-JUN-2000) Botanisches Institut,
JOURNAL Christian-Albrechts-Universität, Olshausenstr. 40, Kiel 24098,
Germany
FEATURES
source
location/Qualifiers
1..1607
/organism="Scenedesmus obliquus"
/strain="wild type D3"
/db_xref="taxon:3088"
<1..1212
/codon_start=1
/evidence=not experimental
/product="Fe-hydrogenase"
/protein_id="AAG59621.1"
/db_xref="GI:12581498"
translation="PHWQTLDELAKPKRKVMIAQIAPAVAGIAETGLNGDVTG
QWVATGRLMIGPDYVFDLFGADLTIMBETLHLLOHLEQHPKKEPLPMFTSCCP
GVAVAKSNPELIPYLSCKSPQMLLAVIKNYFAAGAKPEICNVSVMPCYRKS
GEAEKRSSTHRRAGRDVDHMTTAEIGKI FVEGIGIKLNELOES PPNPVGSGGG
LI FGTGCVMEALPTVEYEVVYALGSGHBSLSTSTWTPQARSPPSPQAPARS
RFLQOTBSGTTLNAVANGNAGNAKLKQLAAGSKYDFTEVMAKPGSGCTGGGQPO
NKKOILQKROAMVDLDERAVIRTEPNLIGALVEKFLGEPNGHAKHBLTHYVAG
VPDRSDEAW"
misc_feature
292..375
/feature="Region: H-cluster 1"
misc_feature
451..486
/feature="Region: H-cluster 2"
misc_feature
952..999
/feature="Region: H-cluster 3"
BASE COUNT 330 a 450 c 520 g 307 t
ORIGIN
Query Match 17.7% Score 424.6 DB 8: Length 1607;
Best Local Similarity 64.2% Pred. No. 1.6e-43;
Matches 816; Conservative 0; Mismatches 359; Indels 96; Gaps 8;

```



```

Oy 603 CTCACGAGCAGCTGAGAGCCCACTCCGACAGCCGCTGCCATGTTACACAGC 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CTTACAGACACCACTGAGAGCAGCAGCCCAAGAGAGAGCCGCTGCCATGTTACACAGC 297
Oy 663 TGTCTGCCCCGGCTGATTCCTATGCTGAGAAATCTTACCAGACTGTATCCCTTACG 722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 TGTCTGCCCCGGCTGATTCCTATGCTGAGAAATCTTACCAGACTGTATCCCTTACG 357
Oy 723 AGCAGCTGAGAGAGCCCAAGATGATGCTGAGAGAGAGAGAGAGAGAGAGAGAG 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417
Oy 783 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 GAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477
Oy 843 CAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
Oy 903 GACCAAGTATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 GACCAAGTATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
Oy 963 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
Oy 1023 CTGTTGCGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 CTGTTGCGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
Oy 1083 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
Oy 1143 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 GGC-----CTGGAGCC 776
Oy 1203 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
Oy 1263 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 --CTTCAAGGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 879
Oy 1323 GCCAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 880 GCCAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
Oy 1383 TAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 940 TAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999
Oy 1443 CGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1000 CAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1054
Oy 1503 AAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 -AACGCGCGGCTGATTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
Oy 1563 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1114 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173
Oy 1623 GGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1174 GGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
DEF07229          5251 bp  DNA      linear  BCT 10-FEB-1996
LOCUS             Desulfovibrio fructosovorans potential NAD-reducing hydrogenase
DEFINITION
ACCESSION         U07229
VERSION           U07229
KEYWORDS          operon genes, complete cds.
SOURCE            U07229.1  GI:466362
ORGANISM          Desulfovibrio fructosovorans.
                  Desulfovibrio fructosovorans
                  Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
                  Desulfovibrio.
REFERENCE
AUTHORS           Malki,S., Salmataime,I., De Luca,G., Rousset,M., Dermoun,Z. and
                  Belach,J.P.
TITLE             Characterization of an operon encoding an NADP-reducing hydrogenase
                  in Desulfovibrio fructosovorans
JOURNAL           J. Bacteriol. 177 (10), 2628-2636 (1995)
MEDLINE           95270577
PUBMED           7751270
REFERENCE         2 (bases 1 to 5251)
AUTHORS           Malki,S.
TITLE             Direct Submission
JOURNAL           Submitted (01-MAR-1994) Souria Malki, Bioenergetique, Ingenierie
                  des Proteines, Centre National de la Recherche Scientifique, 31
                  chemin Joseph Aiguier, Marseille 13402, Cedex 20, France
FEATURES
source            1..5251
                  /organism="Desulfovibrio fructosovorans"
                  /db_xref="taxon:878"
-35 signal
-10 signal
RBS              85..90
                  /note="potential NAD-reducing hydrogenase subunit"
CDS              109..114
                  /codon_start=1
                  /transl_table=11
                  /protein_id="AAA87054.1"
                  /db_xref="GI:466363"
                  /translation="WQNSTCOAVGRCRPERHVLPOPLYREYVQFIESIPQKCHLVT
                  VHKQSVFGYLPISVQFVADHMEVPLAQVGVSTFTFTVWEGKPIVSCMTA
                  CFVKADRVNVAHFKQELKIDIGVTPDRFSIDLRCVGCALAPVIMVGEKRYGNVT
                  PQGVAKIADAEY"
RBS              925..929
                  /note="potential NAD-reducing hydrogenase subunit"
CDS              937..1317
                  /note="potential NAD-reducing hydrogenase subunit"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="AAA87055.1"
                  /db_xref="GI:466364"
                  /translation="WSTIRSFEDLAKROEILDRKAARNGKTIIVSLATCSIAAGK
                  VAMEQDEVANKGLIVGFQMSQCTCYAPVEITLPGQDPVPGVDENRAREL
                  VTEYVMKEPVEVGIIPMYERVVL"
RBS              1330..1334
                  /note="potential NAD-reducing hydrogenase subunit"
CDS              1342..2814
                  /codon_start=1
                  /transl_table=11
                  /protein_id="AAA87056.1"
                  /db_xref="GI:466365"
                  /translation="MAATTEKKOLRIATRNGFIIDPESIDDIYIALRGYEGAKVLTM
                  TPAEVDLVKRSGLRGAGRPDTGIKMGIALGNKADQKVMYCNADGEPREMDRAVL
                  EDDPDSVVENAALIGYRATGATRGTYIRAEYFLAIRKLKALIDDAREYILBENIFGS
                  GFDFDIELKYGAGFVCGEETVALIRSMGKRQEPVTKPFPASQCYWEKPTIVNNET
                  FANIPAIILINGADWFSIGTATSKGTQVFPALAGKIQWGLIIVPMGISLREYIFDIG
                  GCPDGKAFKAVOTGSPSGGALNKDLVAIIVYESIACAKSIMGSGMVMDDDCWVS
                  VAKFPLDFMTDCTGCTPCRIKGRLEVAIIDIRIKGAKTRADLPBLKLSLTKDTA
                  LCGLOTMPNPILSTMDTPANEYEAHVDDKCPAPVCTALLTYTIDPACTCGGLCTR
                  VCPVEICSGTRKHPTIDTRCIKGCACIDKCKPFSIIKO"
RBS              2823..2827
                  /note="potential NAD-reducing hydrogenase subunit"
CDS              2834..4591

```



```

exon      /number=3
           /260..2325
           /gene="hyda"
           /number=4
           /2326..2621
           /gene="hyda"
intron    /number=4
           /2622..3089
           /gene="hyda"
exon      /number=5
           /3090..3400
           /gene="hyda"
intron    /number=5
           /3401..3877
           /gene="hyda"
exon      /number=6
           /3878..4190

BASE COUNT 1002 a 1335 c 1474 g 1190 t
ORIGIN
Query Match 10.2%; Score 244.6; DB 8; Length 5001;
Best Local Similarity 70.1%; Pred. No. 2,1e-21;
Matches 344; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 599 CCGCTCACCGACGACCTGGAGCCCACTCCGACGAGCCGCTGCCATGTTAC 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2612 CTGCTTGACGAGACACCTGGAGCAGCACTCCAAAGAGAGAGCCCTGCTGTTAC 2671

QY 659 CAGCTGCTCCCGGCTGATGCTATGCTGAGAAATCTTACCGGACCTGATCCCTTA 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2672 CAGCTGCTCCCGGCTGATGCTATGCTGAGAAATCTTACCGGACCTGATCCCTTA 2731

QY 719 CGTGAACAGCTGCAAGAGCCCGAGATGATGCTGCGCCATGTCAGTCTTACCTAGC 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2732 CGTGTCTTCTGCAATGCTGCGCCAGATGATGCTGCGCCAGATCATCAAGACTACTTCCG 2791

QY 779 GGAAGAAAGAGGATGCGCCCAAGACATGTCATGCTGTCATATGCCCTGCACGCC 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2792 TGCCTGAGGCGCGGCGCAAGCTGAGGACATCTGCAAGCTGATGCTGCTGCTGCG 2851

QY 839 CAAGCAGTGGAGGCTGACCGGCACTGCTTCTGTGAGACGCCGACCCCTGCGGCA 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2852 CAAGCAGGCGGAGCTGACCGGCACTGCTTCTGTGAGACGCCGACCCCTGCGGCA 2908

QY 899 GCTGACACAGTATCAACCACTGAGCTGGGCAATCTTCAAGAGCGCGCATCA 958
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2909 CGTGAACAGCTGATCAACCACTGAGCTGGGCAATCTTCAAGAGCGCGCATCA 2968

QY 959 CCTGACGAGCTGCGCGGAGCGAGTGGGCAATCTTCAAGAGCGCGCTCGGCGCG 1018
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2969 GCTGAACAGCTGATCAACCACTGAGCTGGGCAATCTTCAAGAGCGCGCTCGGCGCG 3028

QY 1019 CGTGTCTTCTGCAATGCTGCGCCAGATGATGCTGCGCCAGATCATCAAGACT 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3029 CGTGTCTTCTGCAATGCTGCGCCAGATGATGCTGCGCCAGATCATCAAGACT 3088

QY 1079 GTTACGCGCA 1089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3089 GTTACGCTCA 3099

RESULT 9
LOCUS DfrhYDABG 2887 bp DNA linear BCT 30-MAR-1997
DEFINITION D. fructosovorans hyda and hydB genes.
ACCESSION Y11759
VERSION Y11759.1 GI:1914863
KEYWORDS Fe-hydrogenase; hyda gene; hydB gene; iron-only hydrogenase.
SOURCE Desulfovibrio fructosovorans.
ORGANISM Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
Desulfovibrio.
REFERENCE 1 (bases 1 to 2887)
AUTHORS Casalo, L., Hatchikian, E. C., Forget, N., de Philip, P., Belach, J. P.

```

```

TITLE and Rousset, M.
JOURNAL Molecular study and partial purification on iron-only hydrogenase
AUTHORS in Desulfovibrio fructosovorans
TITLE Unpublished
JOURNAL 2 (bases 1 to 2887)
AUTHORS Direct Submission
TITLE Submitted (11-MAR-1997) L. Casalo, CNRS, UPR9036 BIP03, 31, Chemin
Joseph Aiguier, Marseille, 13402 Cedex 09, FRANCE
FEATURES
source location/Qualifiers
1..2887
/organism="Desulfovibrio fructosovorans"
/strain="DSM 3604"
/db_xref="taxon:878"
445..1721
/gene="hyda"
445..449
/gene="hyda"
456..1721
/gene="hyda"
/EC_number="1.12.2.1"
/function="Fe-hydrogenase large subunit"
/codon_start=1
/transl_table=11
/product="cytochrome-c3 hydrogenase"
/protein_id="CAA72423.1"
/db_xref="GI:1914864"
/db_xref="SPTREMBL:O08311"
/translation="MSRIEMAKIFIEQVITPPPGTNDLDAIVYQVDETKICGDTCHGY
CTGAITGESGPHRVVDPACINCGCLTHCPVAIYETVSFVEIEKLIKDKNVK
IAMPAPAVRYALGDPFGMPGLAVTTEHMLTGKQGFNDVWNEFTADVTIIEGSEL
IARITKLDKPLPQFTSCCPGOKYAEIYFELLPHFSSCKSPICGMGLATYGAKE
LGEYKROIYVSIIMPCTAKKPEGMRENDASFRIDATINTRELAVMMKKGIIDLPK
IANGKRDVMEGSTGATIFGSGVMEALAFAYOALTKRPOSMDFAVGLNLIK
EATINIGTIDVAVAVVNGSKNPAKCYDEVKAKSPYHFIETPACFEGCGMGSGPIIP
TVLESNRRITTKFYASLKRRLALYDAORA"
1814..2198
/gene="hydB"
1814..1819
/gene="hydB"
1824..2198
/gene="hydB"
/EC_number="1.12.2.1"
/function="Fe-hydrogenase small subunit"
/codon_start=1
/transl_table=11
/product="cytochrome-c3 hydrogenase"
/protein_id="CAA72424.1"
/db_xref="GI:1914865"
/db_xref="SPTREMBL:O08312"
/translation="MSIILATTRRGFMKTACVLTGGLIIGLRLTSKAVAAKOLXEYIM
DRVNAVYGDADFVKYRASODNAOVITLTKFLHEPLSHSEHLHTTKVDRSKDLAAL
RAKGLYPTPRAKEFEKEPYFQ"
1824..1925
/gene="hydB"

BASE COUNT 613 a 985 c 776 g 513 t
ORIGIN
Query Match 10.1%; Score 242.4; DB 1; Length 2887;
Best Local Similarity 55.4%; Pred. No. 4,5e-21;
Matches 555; Conservative 0; Mismatches 411; Indels 36; Gaps 3;

QY 368 GCGGCTCGCGGAGCTTGCCAGCCCAAGAGACACCCAGCGCCAGCAGCTCGCGTGA 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 GTCTTGTGTCGCGAGATGAGACCAAGCTCAAGGACAAAGACGTCAAGTATCCCAT 769

QY 428 GGTGCTCCGCGGCTTGCTGCTATGTCGAGACCTGGGCTGGCGCGCGCCAGC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 770 GCGGCGCGCGGCGGCTGCTGCTATGTCGAGACCTGGGCTGGCGCGCGCGGT 829

QY 488 CACCCCAAGCAGCTGCGCGGAGGCGCTCGCGCTCGGCTTGAAGAGGTGTTGAC 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 830 CACCAACGACATGCTCAGCGGCGCTCAAGCAGCTTGCGCTTGAACAGTGGAGCA 889

```

```

QY 548 GCTGTTTGGCCGCACTGACCATCATGAGAGGCGACGAGCTGCTGACACCGCTCAC 607.
DB 890 CGAGTTCAACGGCCGACGTCACCATCTGGAGAGAGGCTCCGAGCTTCCGCCCATCAC 949
QY 608 CGAGACCTGGAGAGGCCACCGCACTCCGAGAGAGCGCTGCTCATGTTTACCACTGCTG 667
DB 950 CAAAAGCT-----CGACAAAGCCCTGCGCCCAAGTTCACTCTCTG 991
QY 668 CCGCGGCTGATCGCTATGCTGAGAAATCTTACCAGCACTGATCCCTGCTGAGAG 727
DB 992 TCCCGGCTGGCAGAAAGTACGCGGAGACTTCTACCCGAGCTGCTGCGGACTTCTCTC 1051
QY 728 CTGCAAGAGCCCGCAGATGATGCTGCGCCATGCTCAAGTCTTAAGCGGAGAAAGAA 787
DB 1052 GTGCAAGTGGCCCTACCGCATGATGGTCCCTGCGCAAGACTACGGGGCCAGAGAGCT 1111
QY 788 GGGCATCGCGCCAAAGACATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 847
DB 1112 CGGCTACGAGCCCAAGCAGATCTACCGCTGCTCATGCTGCTGCAAGCGCAAAATTT 1171
QY 848 GAGGCTGACCGCACTGCTTCTGTGAGACGCGGACCCCACTGCGCCAGCTGAGCA 907
DB 1172 CGAAGGCTGCGCGCGGAGAT-----CGACGCGAGGCGCTTCCGCAATCGACGC 1222
QY 908 CGTATCAACCAACCGTGAAGCTGGGCAATCTTCAAGAGCGCGCATCACTGCGCGA 967
DB 1223 CACCATCAACACCGCGAGCTGGCTTACATGATGATAAAAGCGCGCATCGACTGCGGAA 1282
QY 968 GCTGCGCGAGGCGAGTGGGCAATTCATGAGCGTGGGCTCGGCGCGCGCGTGT 1027
DB 1283 GATCGCCCAATGGCAAGGCGGACCGGTCATGAGGAGATCAACGGGCGCGCGCATCTT 1342
QY 1028 CGGACACACCGCGCTGTCAAGAGCGCGCTGCGGACCGGCTTATGAGCTGTTCAAG 1087
DB 1343 CGGCTTTCGCGCGCTCATGAGAGCGCGCTGCGCTTACCAAGCGCTGACCA 1402
QY 1088 CACCGCGCTGCGCGCTGAGCGCTGAGCGAGGCGCGGATGAGACCGCATCAAGAGAC 1147
DB 1403 GAAACCGCGCGAGAGCTGAGCTTCAAGCGCGCTGCGGCTTAAAGGCAAGC 1462
QY 1148 CAACATCAACATGTCGCGCGCGCGCGGTCATGAGAGCTGTCAGAGCACCGCGC 1207
DB 1463 CACCATCAACATG-----GCGGACCGACGTCATGAGTGGCGCTGTCAGAGCGG 1513
QY 1208 CGCGCGCGCGCGCGCGCGCGCGCGGTCATGAGAGCTGTCAGAGCACCGCGC 1267
DB 1514 CAAGAACTTGGCCAAAGTCTGCGACGAGGTGAAGCGCGGCAATCGCTTACCACTTCA 1573
QY 1268 CGCGGCTTACCGACGAGAGCGGCGGCGGCGGATCACTGCGGTGGCGCGTGGCAA 1327
DB 1574 CGAGTTCATGCTGCGCGCGCGGCGGCTGCGTATGAGGCGGCGGCGGCGGCGGCG 1633
QY 1328 CGGCGTGGGCAAGCCCAAGAGCTGATCAACCAAGATGACAGC 1369
DB 1634 CGTCTGATGATCATGAAACCGACCAAGAGTTCTAGC 1675

```

```

RESULT 10
DVUHYDFE 1964 bp DNA linear BCT 26-APR-1993
LOCUS D.vulgaris (Fel)-hydrogenase alpha and beta subunit genes, complete
DEFINITION
M27212.1 GI:145098
VERSION
M27212.1
KEYWORDS
D.vulgaris: oxamizus (strain Monticello) DNA, clone pJ51130.
SOURCE
Desulfovibrio vulgaris
ORGANISM
Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
Desulfovibrio.
REFERENCE
1 (bases 1 to 1664)
AUTHORS
Voordouw,G., Strang,J.D. and Wilson,F.R.
TITLE
Organization of the genes encoding [Fe] hydrogenase in

```

```

JOURNAL Desulfovibrio vulgaris subsp. oxamizus Monticello
MEDLINE J Bacteriol. 171 (7), 3881-3889 (1989)
PUBMED 89291738
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by
G.Voordouw, 07-SEP-1989.
FEATURES
source Location/Qualifiers
1..1964
/organism="Desulfovibrio vulgaris"
/db_xref="taxon:881"
/notes="[Fe]-hydrogenase alpha subunit"
CDS
237..1502
/notes="[Fe]-hydrogenase alpha subunit"
/codon_start=1
/transl_table=1
/protein_id="AAA23373.1"
/db_xref="GI:145099"
/translation="MSRIEMEXIFYEDHAPPEKADPDPLFIQIDESCIGDSCQY
CPRTGAFDPTGDAHKIPHEELCINGGCLTPVGAIVESGWTVEIKIKAKDKV
IAMPAPAVAYALGDAFG.PVGVTGKMFSALEKELFPHQNDNEFTADVTIWEBGTEF
VORLTKLDPKPLPOTSCPGMKHYVESLYPELPHEMSCKSLGMLTAKTYGADR
MKYBRAKYTVSIMPCTAKKYEGMRPOLMSGHDIDTITDRELATVIKAKIDFTK
LEPDGRDILMESSTGATLFGVTGSMELALRYAQAATGKPKESMDFRGVGLQGVK
EATVAVGVDVAVAVHAGRRFDVCELVKAGKAPWHPFERMACPGCGCGGQPVMP
GVLEADRRSTRMYAGLKKRLAMASARA"
1514..1888
/notes="[Fe]-hydrogenase beta subunit"
/codon_start=1
/transl_table=1
/protein_id="AAA23374.1"
/db_xref="GI:145100"
/translation="MQIVNLTRRGLKAACVVTGAGALISIMTKAVAAAKOLDYMM
DRINGVYADAKFPVRSQDNVQYQKLVADLEKPMHSKABQLLHTHVDRSKAIERM
KAQGVPPRAKEFEQNTPEY"
638 a 633 g 284 c
BASE COUNT 409 a
ORIGIN 672 bp upstream of EcorI site.
Query Match 9.3%; Score 222; DB 1; Length 1964;
Best Local Similarity 54.4%; Pred. No. 1.6e-18;
Matches 537; Conservative 0; Mismatches 415; Indels 36; Gaps 3;

```

Db 971 GTTGTG-----GACACGGCCCAAGAGCATGAGCCGACATCGACACCCGCGA 1021
 QY 926 GCTGGCAACATCTTCAAGAGCGCGGCATCAACTGCGAGCTGCCGAGCGAGT 985
 Db 1022 ACTGGCTACATATCAAGAGCGCAAGATCGATTCACCAAGCTGCCGAGCGAAGCG 1081
 QY 986 GGAACAATCAATGGGCGTGGGCTCGGCGCGCGCTGCTGTTGAGCAACAACCGCGGTG 1045
 Db 1082 CGATACCTCGATGGGGAATCCACCGCGCGCGGACCTGTTGGGGTCAACCGCGCGGT 1141
 QY 1046 CATGAGGCGCGCTCGCGACCGGCTTATAGCTGTTCAACCGGACCGCGCTGCCGCGCT 1105
 Db 1142 CATGAGGCGCGCTCGCGACCGGCTTATAGCTGTTCAACCGGACCGCGGAAAGCAT 1201
 QY 1106 GAGCTGAGGAGGAGTGGCGCGGATGAGCGGATCAAGAGCAACATCAATCATGAGTGC 1165
 Db 1202 GAGCTTCAAGAGGCGTGGCTGCTTCAAGCGGCTGAAAGCAACGATGACGTTGAGCGG 1261
 QY 1166 CGCGCCCGGCTCCAGTTTGAAGAGCTGTAAAGCAACCGCGCGCGCGCGCGCGAGCG 1225
 Db 1262 CGTGAAG-----TGAAGGTGGCGCGTGTGCAAGCGGCGCGGCGCTTCAATGAGCT 1312
 QY 1226 CGCGCGCGACGAGACCCCGCGCGCTGCGCTGAGAGCGCGCGCGCTTCAACAGCGA 1285
 Db 1313 GTGCGAAGCTGTGAAGCGCGGCAAGCGCGCGCTGAGTTCATGAGTTCATGCGCTGCC 1372
 QY 1286 GAGCGGACGAGGCGGCGGATCACTGCGCGCTGAGCGCGCGCGCGCGCGCGCGCAACCGCA 1345
 Db 1373 CGCGGCTGTGTGTGCG 1432
 QY 1346 GAAGCTGATCAACCAAGATGAGCGCGCGC 1373
 Db 1433 CGCGAGTGCACCCGATGATACCGCGCGC 1460

RESULT 11

LOCUS

DVHYDG 2067 bp DNA linear BCT 12-SEP-1993
 Desulfovibrio vulgaris gene for 46-Kd hydrogenase.

DEFINITION

VERSION X02416.1 GI:40829

KEYWORDS hydrogenase; inverted repeat.

SOURCE

ORGANISM Desulfovibrio vulgaris.
 Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;

REFERENCE 1 (bases 1 to 2067)

AUTHORS Voordouw, G. and Brenner, S.

TITLE Nucleotide sequence of the gene encoding the hydrogenase from

JOURNAL Desulfovibrio vulgaris (Hildenborough)

MEDLINE Eur J Biochem. 148 (3), 515-520 (1985)

COMMENT

3888621
 The put. transcription terminator (pos. 87-105) may explain the

poor expression of the 46-kda hydrogenase in log. phase E. coli

cells transformed with pHV15.

FEATURES

source

1. 2067
 /organism="Desulfovibrio vulgaris"

/strain="(Hildenborough)"

/db_xref="taxon:881"

42..47
 /note="(weak) -35 region"

67..72
 /note="(weak) -10 region"

87..105
 /note="(weak) -10 region"

87..105
 /note="put. transcription terminator"

87..93
 /note="inverted repeat A"

99..105
 /note="inverted repeat A"

repeat_unit
 /note="inverted repeat A"

repeat_unit
 /note="inverted repeat A"

RBS 153..158
 /note="pot. ribosome binding site"
 CDS 165..1430
 /note="46-kda hydrogenase (aa 1-421)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA26266.1"
 /db_xref="GI:40830"
 /db_xref="SWISS-PROT:P07598"
 /translation="MSRTYMERLEYEMHTPDPKADPDKLHFOVDEAKCLGDTCSOY
 CPTAATFGMEGRPHSIPHEACINSGCCLTTPENAIYEAQSVPEVEKLDKXKC
 IAMPAAVEATLGDAPFVGSVTGKMLALQKGFACWDETFADVTIWESEF
 VERLTKSDMPLEPFTSCCPGQKTAETYPPELLPHFTCKSPIGMIALATYGAER
 MKYDPRQVYTVSIMPDIKAKYEGRLPELSSCMRDI DATLTRELAVMFKKGIIDAK
 LPDKRDSLMGSESTGATIEGVTGGVMEALFAVEAVTGKKPDSMDFKAVGLDGIK
 ETVLEMDRTTTLVAGLKKRLMAASANKA"
 1430..1435
 /note="pot. ribosome binding site"
 CDS 1442..1813
 /note="13.5-kda protein (aa 1-123)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA26267.1"
 /db_xref="GI:40831"
 /db_xref="SWISS-PROT:P07603"
 /translation="MQIASITRGRFLKVCVTTGAALIGIRMTGKAAVAVKIKDYM
 DLINCYGDAKFPBPASODNTOVALYSLYELKPLGHKSHDLHTHMPDKSGVEL
 TTAGKLPNPRASEBEPYE"
 2014..2040
 /note="put. transcription terminator"
 repeat_unit 2014..2025
 /note="imperfect inverted repeat B"
 repeat_unit 2029..2040
 /note="imperfect inverted repeat B"
 BASE COUNT 459 a 688 c 597 g 323 t
 ORIGIN

Query Match 84%; Score 201.4; DB 1; Length 2067;
 Best Local Similarity 53.4%; Pred. No. 5,6e-16;
 Matches 517; Conservative 0; Mismatches 416; Indels 36; Gaps 3;
 QY 432 GCTCCGCGCTTGTGCTGCTATTGCGGAGACCTGGGCTGGCGCGCGCGCACACC 491
 Db 483 GCGCCGCGCGTGTGCTATGACATGCGGCGAGCGCTTGCGGATCGCGGCTCGTCACC 542
 QY 492 CCCAAGACGTGGCGGAGCGCTCGCGCGCTGCGCTTTGACGAGGTGTTGACACGCTG 551
 Db 543 ACCGCGAAGATGCTCGCGCGCTGCGAAGCTCGGCTGCTCATGCTGGGACACCGAG 602
 QY 552 TTGGGCGCGACCTGACATGAGGAGGCGAGGACCTGCTGACCGCGCTCACCGAG 611
 Db 603 TTACCGCTGACGTGACATCTGGGAAGGGGTGCGAGTTCTGGAAAGCGCTCACCAA 661
 QY 612 CACTGGAAGGCGACCGGACCTGCGAGCGCGCTGCCCATGTTACACGAGCTGCGCC 671
 Db 662 -----GAAAGCGACATGCGCGCTGCCGAGTTCACCTGCTGCTGCGCC 704
 QY 672 GGTGATGCTATGCTGGAAGAAATTAAACCGGACCTGATCCCTACGTCAGCACGTGC 721
 Db 705 GGTGACAGAAATATCCAGACCTTACTACCGGAATCTGCGGACCTTCTCCAGGTGC 764
 QY 732 AAGAGCGCGCAGATGATGCTGGGCGCATGCTCAATCTTACTAGCGGAAAGAGGCG 791
 Db 765 AAGTGCCTATGCGCATGAAGCGGCGACCTGCGAAGACCTTACGCGGCGGATGAAG 824
 QY 792 ATGCGCCAAAGACATGCTGATGCTGCTATGCTGCTGCGGACGCGGACGAGTGGAG 851
 Db 825 TAGGACCCCAAGAGGTCTACACCGCTTCATATATCTGCTGATTCGCAAGAGTACGA 884
 QY 852 GCTGACCGGAGATGTTTGTGTGAGCGCGGACCCACCTGCGGACGTGACCAAGCTC 911
 Db 885 GGGTTCGCTCC-----GAAGTGAAGTCCAGCGGATGCGGACATCGACGCGACG 935

Qy	ATCACACCGTGGAGAGCTGGGACCAATCTTTCAAGAGAGCGCGCATTCACACTGGACGACTG	971
Db	936 CTGACACACCGGTGAGCTGAGCCTTACATGATTCAGAGAGCGCGTATTCGACTTCGCAATCTC	995
Qy	972 CCGGAGGGCGAGTGGAGCAATTCAAATGGGCGTGGGCTTCGGGCGCGCGGTGCTGTTGGC	1031
Db	996 CCGGACGGCGAAGCGTGAAGCAGCCTCATGGGTGATTCACCGGCGGGTGCACCATCTTTCGGC	1095
Qy	1032 ACCACCGGCGGATCATGGAGGCGCGCTGGGCAAGCGCTTATGAGCTGTTACAGCGGACAG	1091
Db	1056 GTCAACCGGCGGCGTTCATGAAAGCGGCACTCCGCTTCGCTTACGAAGCCGTATCCGGCAAG	1115
Qy	1092 CCGCTGCGCGCTGAGCCTTAGCCGAGGATGGCGGCGCATGTGACGGCATTCAGAGACCAAC	1151
Db	1116 AAGCCCGACAGCTGGGACTTCAGAGCCGTGGCGGCTTTGATGAGCATTCAGAGAACCAAC	1175
Qy	1152 ATCAACATGATGCCCGCGCGCGCGGTGCAGTTTATGAGAGCTGCTGMAAGCACCAGCGCGCGC	1211
Db	1176 GTCAACGCTG-----GGGATACCGACGTCAAGATCGCGCGTGGTGCACAGGGGACAAAG	1226
Qy	1212 GCGCGCGCGGAGCGCGCGCACCGGCAACCCCGGCGCGCTGGGCTGGGAAGGAGCGCGC	1271
Db	1237 CGGTTCAAGCAGGTCTGCGACGATGTGAAGGCGGGCAAGTGCCTTATCATTTATCGAA	1286
Qy	1272 GCGTTTACACAGAGAGACGCGAAGGGCGGCAATCAACCTGCGCGGTGGCGGTGGGCAAGGAG	1331
Db	1287 TACATGGCGCTGCCCGGGGCGCTGCTCTGTGGCGGCGGTCAAGCCCGTATGCCGGCGTG	1346
Qy	1332 CTGGGCAACGCCAAGAGACTGATTCACCAAGATGCAAGGCGCGGCGAGGCCAATACGATTT	1391
Db	1347 CTCGAAGCGATGAGACCGGACCAACCGCGCTTTACGGGGGCGTGAAGAACGCGCTCGCC	1406
Qy	1392 GTGAGATTC 1400	
Db	1407 ATGGCGAGC 1415	

LOCUS	CFU298227	3265 bp	DNA	linear	PLN 10-JUL-2002
DEFINITION	Chlorella fusca hyda gene for Fe-hydrogenase, exons 1-5.				
ACCESSION	AJ298227				
VERSION	AJ298227.1	GI:21732234			
KEYWORDS	Fe-hydrogenase; hyda gene.				
SOURCE	'Chlorella' fusca.				
ORGANISM	'Chlorella' fusca				
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales; Scenedesmeceae; Scenedesmus.				
AUTHORS	1 Winkler, M., Heil, B., Heil, B. and Happe, T.				
TITLE	Isolation and molecular characterization of the [Fe]-hydrogenase from the unicellular green alga Chlorella fusca				
JOURNAL	Biochim. Biophys. Acta 1576 (3), 330-334 (2002)				
MEDLINE	22079022				
PUBMED	12084580				
REFERENCE	2 (bases 1 to 3265)				
AUTHORS	Happe, T.				
TITLE	Direct Submision				
JOURNAL	Submitted (13-DEC-2000) Happe T., Molekulare Biochemie, Botanisches Institut der Universitaet Bonn, Karlrobert-Kreien-Str.13, Bonn, NRW, 53115, GERMANY				

FEATURES	Location/Qualifiers
source	1..3265 /organism="Chlorella" fusca" /db_xref="taxon:3073" 67..2246
gene	/gene="hyda" join(67..363,523..618,866..1108,1364..1564,1773..2246)
CDS	/gene="hyda" /EC_number="1.18..99.1" /function="hydrogen metabolism" /codon_start=1

	/product="Fe-hydrogenase"
	/protein_id="CAC83290.1"
	/db_xref="GI:21732235"
	/translation="MCPVVASRHAAGRRAHVABAGFTSEDCPPTPPQALPHWOOA
	LLELAKPKESRLMTAQIASAVRAVAIAETIGLABPDVTI GQLVTGLMLGFDLVFDTL
	FSGDLITMEGETIELHARLODHLQHNPKEEPLPMFTSCPGVMAMVERKNTPELYLS
	SCKSPDMMAEVTKYIVAAOQVOVPSDICNSNMPCYARKOEADREWNTPGAIALRD
	VHDVVTTATVGKIPIFERGI KINEL PESFNDRPIEGCGALLFGTGGWEALRTAY
	EVTQFMGRVDVFEEVRGLEGI KEAEITLRKGDDSPPRAPAGDGQGITTLKYIANULS
	GNAKKIKSLSEGKA KPFIEMVACPGECCIGGGQPSRTDKQII QKFOQAATINDERS
	TIRSRSHDNPFIOALYDKFLGANSHKADLLHTHYVAGIPEEK"
	67..129
transit_peptide	/gene="hyda"
	join(130..1363,523..618,866..1108,1364..1564,1773..2243
mat_peptide	/gene="hyda"
	/product="Fe-hydrogenase"
	/EC_number="1.18.99.1"
	<67..363
exon	/gene="hyda"
	/number=1
	364..522
intron	/gene="hyda"
	/number=1
	523..618
exon	/gene="hyda"
	/number=2
	619..865
intron	/gene="hyda"
	/number=2
	866..1108
exon	/gene="hyda"
	/number=3
	1109..1363
intron	/gene="hyda"
	/number=3
	1364..1564
exon	/gene="hyda"
	/number=4
	1565..1772
intron	/gene="hyda"
	/number=4
	1773..>2246
exon	/gene="hyda"
	/number=5
BASE COUNT	730 a 803 c 884 g 848 t
ORIGIN	
Query Match	7.8% Score 186; DB 8; Length 3265;
Best Local Similarity	71.1%; Pred. No. Ac-14;
Matches 246; Conservative	0; Mismatches 100; Indels 0; Gaps 0;
Oy	1294 GGAGCGGCATCACA CTGCCGTGGCCGTGGCCCAACGGCTGGGCAA CGCCAAGA ACTGA 1353
Dd	1897 GGCAGGCGCATCAG CTTCAAGATTGCAGTAGCCATAGTGCTGGCAATCGCA AGAAGCTCA 1956
Oy	1354 TCACC AAGTCAGAGCCC GCCGAGAGCCCAAGTA GCAGCTTT TGTGA GAATCATG GCGTCCCCG 1413
Dd	1957 TC AAGAC CCCTGT CAAGAGGGCC AAAGTA GTATT CATTAAGTC ATGCGATG CCCCTG 2016
Oy	1414 CGGCGTGTGGGGCGGGCGGGCGAGCCCGGCTCCACAGCAAGGCATCAGCGAAGC 1473
Dd	2017 GTGGCTCATTTGGGAGGGCGGTCAAGCCCGCAGTACTCAACAGCAGATCTCCAGAAC 2076
Oy	1474 GGCAGCGGGCGTGTA CAACTTGAAGAGAA GTCCA CGCTGGCGCGCAGCCACGAGAAC 1533
Dd	2077 GCCAGAGGGCATATG TAACCTTGATGAGGAGCA GTAACATCCGCGCAGACCATGATAAC 2136
Oy	1534 CGTCCATCCCGAGCTTTACGACAGCACTCTCGAGAGCCGCTGGGCCCAAGGGCGACG 1593
Dd	2137 CATTCATCCAGGGCGCTGTAGAAAGATTCTTAGCGCCACCAAGCCCAAGGACATG 2196
Oy	1594 AGCTGTCGACACCCCATAGTGGCGGGGGCGGTGAGGAAGAAAGA 1639

Db 2197 ATCTGTCGACACACTATGTGGCAGGTGGAATTCAGAGAGAA 2242

RESULT 13
AF120457 2585 bp DNA linear BCT 25-APR-2000
LOCUS Megasphaera elsdenii putative ABC transporter (abc) gene, partial
DEFINITION cds; and Fe-only hydrogenase (hyda) gene, complete cds.
ACCESSION AF120457
VERSION AF120457.1 GI:6650983
KEYWORDS
SOURCE Megasphaera elsdenii.
ORGANISM Megasphaera elsdenii.
Bacteria; Firmicutes; Clostridia; Clostridiales;
Acidimicrococaceae; Megasphaera.
REFERENCE 1 (bases 1 to 2585)
AUTHORS Alta, M. and Meyer, J.
TITLE Characterization of the gene encoding the [Fe]-hydrogenase from
Megasphaera elsdenii
JOURNAL Biochim. Biophys. Acta 1476 (2), 368-371 (2000)
MEDLINE 20135753
PUBMED 10669801
2 (bases 1 to 2585)
REFERENCE Alta, M. and Meyer, J.
AUTHORS Direct Submission
TITLE Submitted (15-JAN-1999) Departement de Biologie Molculaire et
JOURNAL Structurale, CEA-Grenoble, 17 rue des Martyrs, Grenoble 38054,
France

FEATURES
source . location/Qualifiers
1..2585
/organism="Megasphaera elsdenii"
/strain="ATCC25940"
/db_xref="ATCC:25940"
/db_xref="taxon:907"
1..817
/gene="abc"
1..817
/gene="abc"
/codon_start=2
/transl_table=11
/product="putative ABC transporter"
/protein_id="AA22113.1"
/translation="STLKVLAGOLAPITAGTITIGQTVAFGFPTOOLPQDEGMVID
VYQEHGPTNPGFORISATRLDQFLFTBEMQYTIATLGGERRRLYLKLMQDR
NYLLDERPNDIDPILTYLERYLDFTQGVVNVSGDRYFLDRVYDKLPTLEKRWDR
FYGDSEYLEKTYSHQEKGEKATAPSLTKTNITPAKTAQLTSROSELRITIEL
PRYEMLKGIINAAIAAGSNVEVSLSEQKEKVDALTERWCELEDLKA"
stem_loop 829..869
gene 998..2466
RBS /gene="hyda"
998..1002
CDS /gene="hyda"
1012..2466
/codon_start=1
/transl_table=11
/product="Fe-only hydrogenase"
/protein_id="AA22114.1"
/translation="MEPFHREFEKIDRPVIDEHNCAVOVDYTKCKNCTLCRRACADT
QVLVYSLSTGDMPICVHCGSSACPGFAIVEVDVAKAALKDEKIVIPOTA
PAVRVGLGAEFGMDPGTFVEGKVAALRTLDADYVDDPGADLTMEATLRLHLO
SEELIPORTSCPAWVERFAETFFPDILQHSSTSPISLIPVITPAADKNIDPK
KIVNVCVPTCAKKAIRREPELSASGLFMDEPEIRDTDICTTRELAMQIDENIDFA
SLDESKFDPAFGASGGRIFGNSGVMAAIRTAHMTGSRPAPDFTPEPVAGLO
GKAKATVIGFVHLVAAISGLGNARAFIDDLIKDAFEDYSFIEVMACPGGICGGG
QKRVKLPQVKYQEARTAISYKSDDEETDIKASMONPEIRLYEAFHDEPLSMAEFTL
HTYFSDSKDOLGRMKULQTNPMSPKYKPRTEE"
stem_loop 2502..2528
BASE COUNT 663 a 740 c 657 g 525 t
ORIGIN

Query Match 7.5%; Score 180.8; DB 1; Length 2585;
Best Local Similarity 56.6%; Pred. No. 1.9e-13;
Matches 397; Conservative 0; Mismatches 272; Indels 33; Gaps 2;

QY 411 AAGCAGCTGCTGTCAGAGTGTCTCCGCGGTGCTGTGCTATGTGCGGAGACCTGAGC 470
DB 1294 AAGATGTCATCTTCCACACGCTCCGCTGCTGCGGTGCTATGAGGACCTTCGCG 1353
QY 471 CTGGCGCGCGCGCGCACACCCCAAGCAGCTGCGCGGAGGCTCCGCGCTTCGCTTT 530
DB 1354 ATGACACCCCGTACCTTGTGAGAGAAAGATGTCGACGCTTCGTAACCTGGTGG 1413
QY 531 GACGAGTGTGTGACACGCTGTTGGCGCCGACCTGACCATCATGAGAGGACGAG 590
DB 1414 GACTATGCTTCTGATACCGATTCGATTCGACCATCATGATGAGAGGACGAG 1473
QY 591 CTGCTGACCGCTTACCGAGCAGCCTGAGAGGCCACCGGCACTCCGAGACCGCTGCC 650
DB 1474 CTCTGACACCGCTGCAATCGA-----AGAAATCCTTATTCG 1512
QY 651 ATGTCACAGCTGCTGCGCGCTGATCGCTATGCTGAGAAATCTTACCGGACCTG 710
DB 1513 CAGTTCACAGCTGCTGCGCGGATGAGTTCGAATTTGGGAACTTCTATCGGACCTG 1572
QY 711 ATCCCTACGTGACGAGCTGACAGAGCCCGACATGATGCTGGCGGCAATGTCAGTCC 770
DB 1573 CTCCAGCATCTGTCTTTCAGCAGAGGCCCATCATGATCTTCAGCCCGGTATCAAGAC 1632
QY 771 TACCTAGGAGAAAGAGGATCGCGCAAGAGCATGCTATGCTATGCTATGCTCC 830
DB 1633 TATTTGCGCGAGCAAAACATGACCCCAAGAAATGTCMACTGCTTACGCGCG 1692
QY 831 TGCAGCGCCAGCAGCTCGAGGCTGACC-----GCGACTGCTTCTGTGAC 878
DB 1693 TGACGCGGAGAAAGCCGAAATCGCGCTCTGTAAGTGAAGCTTCGCGCTTCTGCG 1752
QY 879 GCCGACCCACCTCGCGCAGCTGACACGATCATCACCGGTGAGCTGGCAACATC 938
DB 1753 GATGAACCGGAATCGGATACGATATCTGATCATCAGACCGGGAATGCGCCAGTGG 1812
QY 939 TTCAAGAGACCGGATCATCTGCGCGAGCTGCGCGGAGGAGTGGACATCAATG 998
DB 1813 ATCCAGGAGCAAAATGACTGACTTTCGAGCTCGAAGATAGCAATTCGACAGGCTTTT 1872
QY 999 GCGCGGCTGCGCGCGCGCGCTGCTTTCGACACACCGGCGGTGATGAGGCGGCG 1058
DB 1873 GCGGAGCTTCGCGCGCGCGCGTATCTTCGCAACTCCGGCGGTGATGAGCGGCT 1932
QY 1059 CTGGCAGCGGCTTATGACTGTTCAAGGAGCAAGCGGCTGCGG 1100
DB 1933 ATCCGACAGGATACCATGTTTCAAGCGGCTCGGCGCGG 1974

RESULT 14
AE001705 15158 bp DNA linear BCT 02-JUN-1999
LOCUS Thermotoga maritima section 17 of 136 of the complete genome.
DEFINITION AE001705 AE000512
ACCESSION AE001705
VERSION AE001705.1 GI:4980694
KEYWORDS
SOURCE Thermotoga maritima.
ORGANISM Thermotoga maritima.
Bacteria; Thermocogae; Thermocogales; Thermocogaceae;
Thermocoga.
REFERENCE 1 (bases 1 to 15158)
AUTHORS Nelson, K.E., Clayton, R.A., Gill, S.R., Ginn, M.L., Dodson, R.J.,
Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utecherback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cocton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
Eisen, J.A., Fraser, C.M. et al.
TITLE Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima

JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
PubMed 10360371
REFERENCE 2 (bases 1 to 15158)
AUTHORS Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, R.J.,
Hait, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., McQuinn, K.A.,
McDonald, L., Utecht, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE Direct Substitution
JOURNAL Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source location/Qualifiers
1. 15158
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
139. 2601
/gene="TM0198"
139. 2601
/note="similar to PID:1001492 percent identity: 74.36;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease, ATPase subunit"
/protein_id="AAD35290.1"
/db_xref="GI:4980695"
/translation="MFDKFSKTAQIFVTAQAEAKELGHSYVGTETLILALIKVDRSP
AVLELEMGASYSKREISIMVGMKGFVSPOMTPRAKVTETELAEBAKILGSDK
INPEHLILGILREGEIAHILRLKGLVDITLREIIDIISYSGNKLREBEEDYTA
RSVQLEGFVNLTLELAKKELDVGREIEIRYMOVLVRKKNPVLIGDGVKCT
AIVEGAQRIVAGVPELILKNKVFISIDALVACVKGFEKRRKKILOITTKKX
IILPIDHITVAGSGEADANILKPLAKGEISCGATTPDEIRYIEADALE
RRQKIVKPTSEETLEIKGLKRYEAHKVITYDKALAAVLSKYITDYALP
KAIDVIDEAGARARLVLPPLKMKLELERISDKELAVLQDYEAQAKKEEM
ELAEYKRAEYRRRAETAVVAVDVDAVEVSVTGVPLKIEETEVEKLLNEEA
LHORIVADEPAIKAVARISARGLDPRPIGVGLPTGVGTAKLAEYLF
GDERALIRFPMSEYMERFVSRLIGAPGVYEGEGTLTEKRRRRPPEVILPDLREK
AHPDVFNILLQIMDDGRLTDSQREVDFTITIMTSNIGSSYINSKRTLGPVGN
EKEFEFKIDLVLEVKRTTRPFLARIDETIIFHLKKBHLEQIIDILRLDRKLS
EKMKVILVTSAKAFLEKGFDPVAGARPLKAIQRYVEDPLSEILRNGKFNEDTIV
CRAKDLARTRTGERKEKRVVQ"
2598. 3920
/gene="TM0199"
2598. 3920
/note="similar to SP:P24554 GB:X63155 PID:537229
PID:581233 GB:U00096 percent identity: 63.78; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA repair protein"
/protein_id="AAD35291.1"
/db_xref="GI:4980696"
/translation="MKKVCSCNGVSPKMGRCPCGCEYDTAEVORKNKSGPSL
FLNLEDAGISLERLSTGSELDRAFRGIVGOVLLSGEIGIKSTALQIARPA
ERGLVYVGEESPOQLRADRLILKRLKIDILLENDIDELISLONKRSFVAVD
SIQTVSSDLGSPGISOVKNVTKMTIDPAKRGVPLVILGHVTKGEGIAQKVEH
MDTVAVFEGDRTGRLKIKTKNKGSPDEAVPEIKNGSVQVENSFTEGDLP
GNVLTGVFEKTRFVYQIOLVSKNRTSPKACGVADVNRVMLIAVISKYKLPIE
THDVYVAVGGARIVPADLAIATSISSYLEVSLHNTAAVGEIGLGRVAKVYNIN
RHDVLSKSSGRGIVPIEIEKQGVFEVRLKEAVSIGELIGTPGAD"
3883. 4956
/gene="TM0200"
3883. 4956
/gene="TM0200"
3883. 4956
/note="similar to GB:D26185 SP:P37573 PID:467476
GB:AL009126 percent identity: 65.12; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
gene
CDS
/gene="TM0201"
4975. 6801
/note="similar to GB:U07229 PID:466366 percent identity:
69.31; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADP-reducing hydroxase, subunit D, putative"
/protein_id="AAD35292.1"
/db_xref="GI:4980697"
/translation="MGVKSIVPOELIKIKLISPTGLKALDQIINANFALLFLVD
DKRYEDVIOGFWLDTDSAKLYELSRMDQDAVLSEDTIKIYANVHLVDPTPT
GETGTRHIAERLAQKGVAVVAVSRNRRIISLYKKNVYVNOVDFLSKVTQAIST
LERKQNFENKLLSELEVELENNRVTLAVRLAGFELIRVIERPIVYIGEGR
LARMQRELTEVDVLLVLLIMDYSEEEETAAQNIQDFTREPSISIRVIGY
DYQQAOLDDVLSARGYRLTKVARIPLSISYVNVNMEKTLQJSKASVEDLKKEVG
IGEKRAISISSLKKRKISE"
4975. 6801
/gene="TM0201"
4975. 6801
/note="similar to GB:U07229 PID:466366 percent identity:
69.31; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADP-reducing hydroxase, subunit D, putative"
/protein_id="AAD35293.1"
/db_xref="GI:4980698"
/translation="MRREFKNLRLNSGNETSVRCFALADVTVINGRTIYPDN
LTVIACEAGIEIPALCHHPRLIGESIGACVYEVGARNLQDPACTYKRDGVK
TSDRVKTAARKNALLISERNDCMTANGRCFODLITRYDEPIFGTLEGV
DRSSPAIYRDSKTCRCRCYACSELQGMHTYSKVERGHRTYPTPEMDPVETEDCI
GGQCAACFCFGAIVENSAAVVLSELEKRIILVQYAPRAVIAIGEFYAGDTIS
TGQVVALRLRGFDVDPDTPGADLTIMESSEFLERLEKQDLEDLPMTSCCGWV
LVEKYPELARTLSAKSPQGMILSMVITYPAEKLVKPEDIPIHYSIMPTAKKDEAL
RQMLVNGVPAVDVLTRELGKILRMKIPFANLPEEYVAPIGISGAALALGCTG
GVMEALATVAELKTKGLPKTIVEEVGKGVAEADIDDKKIRIAVHGTAVNRN
LVKELIREVYVHPVEVMAKCGGCTGGGQGVSDPELIRARAAYITIDERMTLRS
HNPFAIKLYEEVLEHPLSHRAHBLHTIYEDRSRKRKLAVK"
6846. 7748
/gene="TM0202"
6846. 7748
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD35294.1"
/db_xref="GI:4980699"
/translation="MKKVTIVVLLTISILFGETLNPGLPALIPVPVINDKIPITV
KIKINKPBEVAKIVSKEDVPAVLPVTVGANLQKGVRIKLVGVHMKVFLVAVSD
ATFGMSLRQGEVYTPHGGQIVDMRFLSAGLITLQDVNLIYAPQEIYALFK
SGKVVYALPPEFVSMCLDRKVVLDQKQWKGELGVPGRIPIAGLTVREGVDKETEY
KVERALDISIRMKENDETVOLESSEKLGIPAKILKSLRIEPEYVPEKCREVET
FLKINELYPEGFEKIPDEGFWK"
7745. 8476
/gene="TM0203"
7745. 8476
/note="similar to PID:2440148 percent identity: 50.22;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein, cystw family"
/protein_id="AAD35295.1"
/db_xref="GI:4980700"
/translation="MRAVFCILVLILAMVLLHFLPSSLLLPGPVETFKVIFMLNRE
TFEALISTLKGAVSTFIVAVGLPVGFPWIGIDRYEPLRPVTVQAAVPSVSWIV
VIFLMGIGMOGQVVISLSLIPVAILFTVGSQNSVVRKLEWKKVYRVRMILKEYY
IGSFMPVVISLIVSNGNWKAVYKMGSLGDSGLVLISMARQYDVPVVALTITFT
VVLGISFERSVKVLAARWKKRRLS"
8452. 9153
/gene="TM0204"
8452. 9153
/note="similar to GB:AE000782 percent identity: 68.20;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"

```
/protein_id="AAD35296.1"
/db_xref="GI:4980701"
/translation="MEVEVILKVEGLRDPNGVKVLENMGFSVGKGERVALLSPSCG
GTTFLRIVSGEDYQKVFPTDKIGYFOEPRLLPMFTTENMLLRDRLASL
LEKVELKGFENHYPMQISEGMQRVNFVALLVDPLLDDPEMLDLKTKMKVSL
LYDLNOKRPFSTVFTHNVKFAVFLSERIFLISGRPSKILDEVKLKAIDFTDELF
RLEKVIERLNLIR"
/complement(9137) .11479)
/gene="TM0205"
/complement(9137) .11479)
/notes="similar to GP:2633959 percent identity: 63.64;
identified by sequence similarity; putative"
/codon_start=1
/translation="
/product="ATP-dependent DNA helicase"
/protein_id="AAD35297.1"
/db_xref="GI:4980702"
/translation="MLCSRYFTSSLFMLGEBALPTLLEBFLNEVEKMLKNQVTRLIHQ
LKELDLPLENDLEKLAQFLDYKEIPLNPEARKRYRIQSLKEMIKRSMFLID
VLESGEVDLSTDIQYAKGVGNPKKKIKGLIETRLDLEFPEDYEDRRKIFLN
DLPEKVTTOGKIVSVETKRONNMLTAVISDGLVHPLKFNODYLOTYKOLTG
KEVFTGYKSNATGYQVEIHNAEVPKSGEVYRRLPIYRLTSGISOKMKRTEPEN
TSLCCSLKETLPERILEKRLGVKQATYGMHPTPTFYHLEKARELVEELVQLQ
AFQKIRKEREGKIGIPK1EGKLAEEFISLPEKLTNAQRAHQEIRNMISEKPMNR
LQGVGSGKTVAQALAILDNYEAGFQTAFTMPTSLIAOHYRITVESFSKNIHVAL
LIGATPSEKIKVQALINDYVIGTALIOEDVHFNKGLVITIDEOHFGVKORE
ALMNGKMDTLVMSATPIPRSMALAFLVDGLDVTVDENPGRKEVOTMLVMDRNE
VIEFPAQEVNRGQAFIVPLIESPKLVAKSVENYELSKVEPFELKLMHGRYS
OEKDVLMLEFAGRIIVLIVSTVIVGVIDVPRANVAVENPERFELADHOLRGVNG
RGQEAIVCFVLVGDVEEMERLRFELTNTDGLKLAELYDLKTRGPGEFVGVKHGSG
FKVADYRLDKLEMARDEVDIEVGEIHELPERIKLEVG"
/complement(11527) .12042)
/gene="TM0206"
/complement(11527) .12042)
/gene="TM0206"

Query Match
Best Local Similarity 54.1%; Score 166.6; DB 1; Length 15158;
Matches 402; Conservative 0; Mismatches 314; Indels 27; Gaps 2;
```

```
QY 891 CTCGCCAGCTGACCAAGCTATCAACACCGTGAAGCTGGGCAACATCTTCAAGAGCC 950
Db 6183 -----GATGATAGTTGTTCTTCAACCAAGAGGAATCTTGGAACTTAAGATGAAG 6234
QY 951 GGATCAACCTGCGGAGCTGCCAGAGGCGAGTGGGCAATCCATAGGCGCTGGGCTGG 1010
Db 6235 AAGATACCGTTTGGCAACCTTCCAGAGGAAGATACAGCGCACCGCTTGAATCTTACG 6294
QY 1011 GGCGCGGCGTCTGTTGGGACACACCGCGGTGCATGAGAGCGCGCTGCGCACGCGCC 1070
Db 6295 GGAGCGGCGGCACTCTTGGCGTTACAGCGGTGTGATGAGGCTGCTTTGAAACTGGG 6354
QY 1071 TATGACTGTTTACGCGGACAGCGCGTCCCGCTGAGCTTGAGGAGGTGCGGCGCATG 1130
Db 6355 TACGAACCTCAAAACGGGAAGGCACTACCGAAGATCGTTTGAAGAGGTGAGAGCACTC 6414
QY 1131 GACGCGATTAAGAGACCAACT 1153
Db 6415 AAGAGTGTCAAGGAGCCGAAT 6437

RESULT 15
DHYDC 1940 bp DNA linear BCT 06-JUN-1991
LOCUS D_vulgaris hydC gene for Hyd gamma periplasmic hydrogenase.
DEFINITION X57838
ACCESSION X57838
VERSION 1
KEYWORDS hydC gene; periplasmic Fe-hydrogenase.
SOURCE Desulfovibrio vulgaris.
ORGANISM Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
Desulfovibrio.
REFERENCE 1 (bases 1 to 1940)
AUTHORS Stokkermans J.P.W.G.
TITLE Direct Submision
JOURNAL Submitted (15-JAN-1991) J.P.W.G. Stokkermans, Agricultural Univ of
Wageningen, Dept of Biochemistry, Dreijenlaan 3, 6703 HA
Wageningen, The Netherlands
2 (bases 1 to 1940)
REFERENCE Stokkermans J., van Dongen W., Kaan A., Van Den Berg W. and
Veeger C.
TITLE Hyd Gamma, a gene from Desulfovibrio vulgaris (Hildenborough)
JOURNAL encodes a polypeptide homologous to the periplasmic hydrogenase
COMMENT FEMS Microbiol. Lett. 58, 217-222 (1989)
FEATURES
source Location/Qualifiers
1..1940
/organism="Desulfovibrio vulgaris"
/strain="Hildenborough"
/db_xref="taxon:881"
37..41
RBS 46..1918
gene /gene="hydC"
46..1918
mRNA <46..1918
/gene="hydC"
prim_transcript <46..1918
/gene="hydC"
46..1866
CDS /gene="hydC"
1 /codon_start=1
/translation="
/product="Hyd gamma"
/protein_id="CAA40970.1"
/db_xref="GI:40828"
/db_xref="SPTREMBL:Q4606"
/translation="NNAFLNGKVRCEPGRITLLEARENGHFIPTLCELADIGHABGT
CRVCLVEIWRKDEAGPOIYTSCTTPVEEGMRIFTPTVEVRMQRLOVELILLADHDHC
AACARHGDELDVNOFGLSTGRHHPYASRTSDVSPSVVDMGKICRLRVA
VCRNOGVNALTAVTNGITETGLRHNSSQASDCVCGCGCTLVCVGLADRDYER
VDIYXDEPITVTFOPAPAVRVGLCEBGLPFGSSVEGVPYALRLHGLGADVLDITFA
ADVLTMBEGTIELORLRGAKLPLTSCCPGVNFAEKHLPIILPHVSTTRSPQOLG
ALAKYTLARTNVAVERMRVSLMPCTKAKEEAARPEFRDSDVVDVAVLTTRERARL
LRREGIDLAGLEPSPDDPLMGRATGAAVIFCTTGVEAALRTVYHVLNGELAPVE
```

LHAIKYEINVEAVPILEGNSVKVAVNHGLKARQWVEAVLAGKADHVEVMACP
GGCMGGQPSKRAVNPNAQARAALFSLDAENALROSHNPILIGVYSFLGEPCS
NLSHRLHTRYGDRKSEVAAYTMRIDIMHEMTLIGRIVRBDSD"
1892..1918
/gene="hycd"

BASE COUNT 343 a 597 c 641 g 359 t
ORIGIN

Query Match 6.7%; Score 160; DB 1; Length 1940;
Best Local Similarity 52.3%; Pred. No. 7,3e-11;
Matches 451; Conservative 0; Mismatches 375; Indels 36; Gaps 3;

```
QY 374 GCGGAGCTTGGCAAGCCCAAGAGACCCCAAGGCAAGCAGTCTGCGTGCAGGTGC 433.  
DB GGTGAGAGGTGTCAATCGACTATCTTACGACCCCGAAATCGTCAACCGTGTCCAGTTGCG 749  
QY 434 TCCGCGCGTGTGTGCTATTTGCGAGACCTTGAGCGCGCGCGCGCGCGCACACCC 493  
DB 750 CCGGCGGTGGCGGTGGCGCTCGGTGAGAGATTGCGGTGCTCCGCTTCAAGCGTGA 809  
QY 494 CAAGCAGTGGCCGAGGGCTTCGCGCGCTTGCGCTTGAAGAGGTGTTGACACGCTGTT 553  
DB 810 AGGCGAGGTGCCACGCGCTTGCGCTTCTCGGGCAGACGTGTACTCGATACCACTT 869  
QY 554 TGGCGCGGACCTGACCATATGAGAGAGGAGGAGAGTGTGTGACCGCTCACCGAGCA 613  
DB 870 CGACGCGACCTGCTCATATGAGAGAGGAGCAGCACTCTGCAACGCTCT----- 920  
QY 614 CTTGAGAGCCACCCGACTCCGACGAGCCGCTCCCATGTTTCAACAGTGTGCGCGG 673  
DB 921 -----TGGGGCGGGGGGGAAGCTGCCCTCTTCACTCTGCTGCGCGG 965  
QY 674 CTGATGCTATGCTGAGAAATTTAACCGGACCTGATCCCTTACGTGAGCAGCTGCAA 733  
DB 966 CTGGGTAAATTTCCCGAAGACACTTCCCGACATCTCCGATGTCTGACCAACG 1025  
QY 734 GAGCCCCAGATGATGCGGCGCATGTCAAGTCTTACCTAGCGGAAGAAGGCGAT 793  
DB 1026 CTGCGCTCAGCAGTGCCTTGGCGCATTTGCCAAGACTTATCTTGGCGGCACATGAACGT 1085  
QY 794 CGCGCCAAAGACATGTCTATGTGTCATATGCCCTGACGCGCAAGCAGTGGAGGC 853  
DB 1086 CGCACCGAGAGATGCGCGTCTGATGCTGATGCCCTGCACGCGAAGAGAGAGGC 1145  
QY 854 TGACCGGACCTGCTGTGTGTGAGACGCGGACCCACCTGCGCAGCTGACCACTCAT 913  
DB 1146 CGCACGCGCGAATTTCAGGCGGACGCTG-----TCCGGATGTGACGCGAGTGTCT 1196  
QY 914 CACCAACGTGAGAGTGGGCAACATCTTCAAGAGCGCGGATCAACCTGCGGAGCT--- 970  
DB 1197 CACCAACGTGAGTGTGCGCGTCTTCTCCGCGTGAAGGCAATAGACTTCGCGGACTCGA 1256  
QY 971 GCGCGAGGCGAATGGGACATCCATGGGCTGGGCTGGGCGCGCGCTGTGTTGG 1030  
DB 1257 ACCCTGCGCTTGACACACCCCTGATGGGCGGCAACCGGCGCGCTGTCTTTCGG 1316  
QY 1031 CACCAACGCGGTGTCTATGAGAGCGGCGCTGCGGCTATGAGCTGTTCACGCGGAC 1090  
DB 1317 TAGCAGAGGCGGGGTATGAGAGCGGACATGCTTACCATGTGTGAAAGCGCAA 1376  
QY 1091 GCGGCTGCGCGCTTGAAGCTTGAAGAGTGCAGGATGACGCGATCAAGGAGACCAA 1150  
DB 1377 GGAATCTGCGCCAGTAGAATCTGATGCCCTGCGCGATAGAGAACGTGTGAGGCTGT 1436  
QY 1151 CATACCATGTGTCGCGCGCGGTTCGAAGTTTGAAGAGCTGCTGAAGACCGCGCGCG 1210  
DB 1437 GGTCCGCTTGTGAGAGGTAAAGGTTCTCGTGAAGTGCCTGTGTGCATGAGGCTCAAGGC 1496  
QY 1211 CGCGCGCGCGAGGCGCGCGG 1232  
DB 1497 TGCCCGGCAAGATGTGAGGCG 1518
```

Search completed: June 8, 2003, 11:53:21
Job time: 6045 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2003, 09:13:01 ; Search time 526 Seconds
(without alignments)
10270.996 Million cell updates/sec

Title: AJ012098
Perfect score: 2399
Sequence: 1 ACTTACATGAACACACAA.....TGTAAACGACACTTCTGAG 2399

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	3.9	1737	20	AAZ25199
2	83.2	3.5	114955	20	AA53491
3	75	3.1	109519	22	AA508693
4	72.6	3.0	114955	20	AA53491
5	71.4	3.0	1552	22	AAH13714
6	71.4	3.0	1577	22	AAH27127
7	70.2	2.9	65140	22	AAH17184
8	70.2	2.9	125401	22	AAH17186
9	69.8	2.9	1836	21	AAH70438

10	69.8	2.9	1944	21	AAA70434
11	69.8	2.9	1944	21	AAA70435
12	69.8	2.9	1944	21	AAA70437
13	69.8	2.9	2025	21	AAA70433
14	69.8	2.9	2025	21	AAA70436
15	69.8	2.9	2310	21	AAA51944
16	69.8	2.9	2316	21	AAA51945
17	69.8	2.9	2316	21	AAA51949
18	69.8	2.9	2322	21	AAA51946
19	69.8	2.9	2328	21	AAA51947
20	69.8	2.9	2334	21	AAA51948
21	69.8	2.9	2358	21	AAA70447
22	69.8	2.9	2358	21	AAA51966
23	69.8	2.9	2466	21	AAA70446
24	69.8	2.9	2547	21	AAA70445
25	69.8	2.9	4608	21	AAA70473
26	69.8	2.9	4689	21	AAA70471
27	69.8	2.9	27541	22	AAH17185
28	68.4	2.9	3009	24	ABL40024
29	68.2	2.8	53789	19	AAV21187
30	67.2	2.8	1944	21	AAA51613
31	67.2	2.8	2101	22	AAH24245
32	67.2	2.8	2113	22	AAH24245
33	67.2	2.8	2466	21	AAA51614
34	67.2	2.8	2547	21	AAA51615
35	67.2	2.8	2550	24	ABL39955
36	65.4	2.7	44377	18	AAH78508
37	65.4	2.7	44377	18	AAH80414
38	64.6	2.7	50937	21	AAH09469
39	63.8	2.7	3465	12	AAQ14478
40	63.8	2.7	3993	12	AAQ15153
41	63.6	2.7	2322	21	AAA51959
42	63.6	2.7	2341	21	AAA51951
43	63.6	2.7	3015	24	ABL39983
44	63.4	2.6	1182	22	AAH14507
45	63.4	2.6	1248	19	AAV11297

ALIGNMENTS

RESULT 1
AAZ25199 standard; DNA; 1737 BP.
XX AAZ25199;
AC 13-DEC-1999 (first entry)
XX
DT 13-DEC-1999 (first entry)
XX
DE Clostridium pasteurianum hydrogenase #2.
XX
KM Clostridium; hydrogenase; blue-green alga; hydrogen; microbe;
KW Shine-Dalgarno; ss.
XX
OS Clostridium pasteurianum.
XX
PN JP11253166-A.
XX
PD 21-SEP-1999.
XX
PF 13-MAR-1996; 98UP-0062601.
XX
PR 13-MAR-1996; 98UP-0062601.
XX
(AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
XX
DR WPI; 1999-583699/50.
XX
PT Preparation of heterologous hydrogenase in blue-green algae - useful
for production of heterologous and is environmentally friendly
XX
XX Example 4; Fig 10; 12pp; Japanese.

XX A method has been developed for the preparation of hydrogenase in which
CC a blue-green alga introduced by a recombinant vector containing a
CC heterologous hydrogenase gene is cultured in a medium to form and
CC accumulate hydrogenase in the microbe body. Also claimed are: (1) a
CC blue-green alga introduced by a recombinant vector containing a gene
CC having a Shine-Dalgarno sequence consisting of AAGGAA upstream of the
CC hydrogenase structural gene derived from a Clostridium genus microbe;
CC (2) a method for the production of hydrogen by culturing the above
CC blue-green alga under light irradiation; (3) a method for the production
CC of hydrogen by culturing the above blue-green alga anaerobically and
CC then under light irradiation; (4) a method for the production of
CC hydrogen by reacting the above blue-green alga with an electron
CC receptor; and (5) a method for the production of hydrogen by culturing
CC the above blue-green alga anaerobically and then reacting it with an
CC electron receptor. The method is mild to environment. The present
CC sequence represents a Clostridium pasteurianum hydrogenase nucleotide
CC sequence from an example from the present invention.

XX Sequence 1737 BP; 565 A; 363 C; 373 G; 436 T; 0 other;

Query Match 3.9%; Score 93; DB 20; Length 1737;

Best Local Similarity 48.1%; Pred. No. 1.1e-07;

Matches 359; Conservative 0; Mismatches 355; Indels 33; Gaps 2;

QY 398 CGACCCCGCGGACGCTGCGTGCAGTGGCTCCGCGCTGCTGCTATTC 457
DB 660 CAATGCCCTGAAACATGTGATCGCTGCTCCAGCTCCGCGTGCATCGG 719
QY 458 CGAGACCCCTGGCGCGCGCGCGCGCACCCCGCAACGACCTGGCGAGGCGCTCCG 517
DB 720 TGAACCTTTTAATATGAGATTTGCGCTTCACGTACAGAAAATTATAGCGCTTCGCG 779
QY 518 CCGCCTCGCTTTGACGAGGTGTTGACACGCTGTTGGCGCGACCTGACATCATGA 577
DB 780 CCAGCTTGATGATTAATATGTTGATATCACTTCGAGAGCATATGACATTTATGA 839
QY 578 GAGGGGACGAGCTGTGACCGCTCACCAGACCTTGAGAGCCACCGGACATCCGA 637
DB 840 AGAGGCTACAGAACTGTTCAACGCAT-----GGAGATAA 875
QY 638 CGAGCGCGCTGCGATGTTACAGAGCTGCGCGCGCTGATGCTATGAGAAATC 697
DB 876 TGAACCTTTCCATGTTTCAAGCTGCTCCCGGTTGGTGGCGGACCTGAAATTA 935
QY 698 TTACCGGACCTGATCCCTACGTGAGCAGCTCAAGAGCCCGCAGATGCTGGCGGC 757
DB 936 TTATCTGAACTCTTAATATATCTTTCATGACGTTAATCACTCAACAATTTTGGTAC 995
QY 758 CATGTCAAGTCTTACCTAGCGGAAAGAGGCGATGCGCGCAAGACATGCTATGCT 817
DB 996 CGGTAGTAAACCTTATATCTTATCTGACATCAGCGGTCTTGAGCCCAAGAAATGTTTACCGT 1055
QY 818 GTCCATCATGCTCTGACGCGGACAGCTGAGAGCTGACCGGACATGTTTGTGTGA 877
DB 1056 GAAGTATATGCTCTGATCTCAAAAAATTGAAGCAGATCGCCCCCAATGAAAAAGA 1115
QY 878 CGCGGACCCGACCTGCGCGAGCTGACACAGTATCAACGACCGTGGAGCTGGCAACAT 937
DB 1116 CG-----GCTACGCGATATCGATGCTGTTATACACACCGAGAACTGGCAAAAAAT 1166
QY 938 CTTCAAGAGAGCGGCGATCAACTGCGCGAGCTGCCGAGAGGAGATGGAGCAATCCAT 997
DB 1167 GATTAAGATGCTAAATATCCCTTTGTAACTTGAAGATAGGAAACACCTCTGAT 1226
QY 998 GGGCGTGGGCTCGGGCGCGCTGCTTCCGACACCGCGGAGTCTCATGAGGCGCG 1057
DB 1227 GGGAGAAATACAGGGGCTGCTGCTCATCTTTGCTGACAGCGGGAGATTATGAAAGCAGC 1286
QY 1058 GCTGCGACGCGCTATGAGCTGTTACGAGGACGCGCGCTGCGCGCTGAGCTGAGCGA 1117
DB 1287 TCTGCGAGAGTGAAGAACTTTGCTGAAACGCTGAACCTTGAAGATATCATATATAGCA 1346

QY 1118 GGTGCGCGCATGACGCGCATCAAGCA 1144

DB 1347 AGTTCGCGGACTGAATGATATCAAGA 1373

RESULT 2

AA53491/C

ID AAX53491 standard; DNA; 114955 BP.

AC AAX53491;

XX 05-JUL-1999 (first entry)

DE Human adenosine A1 receptor antisense oligonucleotide fragment.

XX Antisense oligonucleotide; multiple target; antisense treatment;

KM impaired respiration; inflammation; lung disease;

KM pulmonary vasoconstriction; inflammation; allergic rhinitis;

KM acute asthma; allergy; asthma; pain; cystic fibrosis;

KM respiratory distress syndrome; pain; cystic fibrosis;

KM pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KM chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KM colon cancer; breast cancer; lung cancer; pancreatic cancer;

KM hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;

KM prostate cancer; 86.

XX Synthetic.

OS MO9913886-A1.

PN 25-MAR-1999.

PD 17-SEP-1998; 98WO-US19419.

PF 09-JUN-1998; 98US-0093972.

PR 17-SEP-1997; 97US-0059160.

XX (UYEC-) UNIV EAST CAROLINA.

PA NYCE JW;

PI WPI; 1999-229400/19.

DR New antisense oligonucleotides used in treatment of, e.g. pulmonary

XX vasoconstriction

PT Disclosure; Page 37; 120pp; English.

PS The specification describes antisense oligonucleotides (AAX52869-X55271)

XX directed against at least 2 mRNAs selected from target genes, coding and

CC non-coding regions of RNAs corresponding to target genes, gene

CC initiation codons, genomic flanking regions, intron-exon borders, the

CC 5'-end, the 3'-end and the junction-section between coding and non-coding

CC regions and all segments of RNAs encoding proteins associated with one

CC or more diseases, conditions or mixtures. The antisense oligonucleotides

CC may be derived from sequences AAX5272-74. These multiple target

CC oligonucleotides (specifically AAX5180-211) can be used for the

CC antisense treatment of diseases and conditions. Typical diseases and

CC conditions are those associated with impaired respiration and
CC inflammation, including lung diseases, pulmonary vasoconstriction,
CC inflammation, allergic rhinitis, acute asthma, allergy, asthma, impaired
CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC hepatic metastases, as well as all types of cancers which may metastasize
CC or have metastasized to the lungs, including breast and prostate cancer.

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 3.5%; Score 83.2; DB 20; Length 114955;

Best Local Similarity 31.5%; Pred. No. 7.6e-06;


```
FT CDS 8342..9364  
FT /tag= m  
FT /product= "EvdH"  
FT RBS 8333..8336  
FT /tag= n  
FT CDS complement (9463..10224)  
FT /tag= o  
FT /product= "EvdI"  
FT RBS complement (10232..10235)  
FT /tag= p  
FT CDS 10424..11176  
FT /tag= q  
FT /product= "EvdJ"  
FT 12027..12455  
FT /tag= r  
FT /product= "EvdK"  
FT /partial  
FT /note= "No start codon"  
FT CDS complement (12108..13022)  
FT /tag= b  
FT /product= "EvdL"  
FT RBS complement (13027..13030)  
FT /tag= c  
FT CDS complement (14410..15363)  
FT /tag= u  
FT /product= "EvdA"  
FT RBS complement (15369..15373)  
FT /tag= v  
FT CDS complement (15380..16414)  
FT /tag= w  
FT /product= "EvdB"  
FT complement 16419..17873  
FT /tag= x  
FT /product= "EvdC"  
FT CDS complement (17870..18934)  
FT /tag= y  
FT /product= "EvdD"  
FT 18374..20906  
FT /tag= z  
FT /product= "EvdE"  
FT CDS 21064..22542  
FT /tag= aa  
FT /product= "EvdF"  
FT RBS 21056..22542  
FT /tag= ab  
FT 22748..24172  
FT /tag= ac  
FT /product= "EvdG"  
FT RBS 22736..22740  
FT /tag= ad  
FT complement (24177..25223)  
FT /tag= ae  
FT /product= "EvdH"  
FT RBS complement (25230..25233)  
FT /tag= af  
FT 25550..26626  
FT /tag= ag  
FT /product= "EvdI"  
FT CDS 26685..30479  
FT /tag= ah  
FT /product= "EvdJ"  
FT 26672..26676  
FT /tag= ai  
FT complement (30557..31876)  
FT /tag= aj  
FT /product= "EvdK"  
FT RBS complement (31885..31888)  
FT /tag= ak  
FT CDS complement (31941..32882)  
FT /tag= al  
FT /product= "EvdL"  
FT complement (33167..34405)  
FT /tag= am  
  
FT RBS /product= "EvdM"  
FT complement (34414..34418)  
FT /tag= an  
FT CDS complement (34449..35210)  
FT /tag= ao  
FT /product= "EvdN"  
FT RBS complement (35219..35221)  
FT /tag= ap  
FT CDS complement (35294..36238)  
FT /tag= aq  
FT /product= "EvdO"  
FT CDS complement (36235..36963)  
FT /tag= ar  
FT /product= "EvdP"  
FT CDS complement (36998..38026)  
FT /tag= as  
FT /product= "EvdQ"  
FT CDS complement (38072..38566)  
FT /tag= at  
FT /product= "EvdR"  
FT CDS complement (38892..40163)  
FT /tag= au  
FT /product= "EvdS"  
FT CDS complement (40216..40890)  
FT /tag= av  
FT /product= "EvdT"  
FT CDS complement (40899..40902)  
FT /tag= aw  
FT CDS complement (40887..41576)  
FT /tag= ax  
FT /product= "EvdU"  
FT CDS complement (41679..42707)  
FT /tag= ay  
FT /product= "EvdV"  
FT RBS complement (42714..42717)  
FT /tag= az  
FT CDS complement (42810..43799)  
FT /tag= ba  
FT /product= "EvdW"  
FT RBS complement (43807..43811)  
FT /tag= bb  
FT CDS complement (43799..44866)  
FT /tag= bc  
FT /product= "EvdX"  
FT CDS complement (45014..45760)  
FT /tag= bd  
FT /product= "EvdY"  
FT RBS complement (45767..45770)  
FT /tag= be  
FT CDS complement (45962..46714)  
FT /tag= bf  
FT /product= "EvdZ"  
FT RBS complement (45952..45956)  
FT /tag= bg  
FT CDS complement (47156..49234)  
FT /tag= bh  
FT /product= "EvdA"  
FT CDS 51627..52715  
FT /tag= bi  
FT /product= "EvdB"  
FT 51620..51622  
FT /tag= bj  
FT 52889..53557  
FT /tag= bk  
FT /product= "EvdC"  
FT CDS 53554..54207  
FT /tag= bl  
FT /product= "EvdA"  
FT CDS complement (54362..55117)  
FT /tag= bm  
FT /product= "EvdB"  
FT complement (55125..55128)  
FT /tag= bn
```



```

FT CDS complement (55135..56094)
FT FT /tag= do
FT FT /product= "EybC"
FT RBS complement (56100..56103)
FT FT /tag= bp
FT CDS complement (56184..56813)
FT FT /tag= bq
FT FT /product= "EybC2"
FT CDS 56961..58709

Query Match 3.1%; Score 75; DB 22; Length 109519;
Best Local Similarity 45.2%; Pred. No. 0.00022;
Matches 463; Conservative 0; Mismatches 535; Indels 27; Gaps 4;

QY 320 CGCGGTGCGCGACCGCGCTGCGAGGCGCTTTGAGTCATGTCCAGAGCGCGCTCCGCGA 379
DB 78769 CGCGGCGGTGCTATCAAGAGCCATGGCGGGGACACACACTGACTCGACACAGCGCGCA 78828
QY 380 GCTTGCCAAAGCCCAAGAGGACCCCAAGCGGCAAGCAAGTCTGCGTGAGGTGCTCCGCGC 439
DB 78829 TCTGAGACGACTTCGCGACCGGCGCGGCGGTGCTGACGACGACCCGACCGGCGAC 78888
QY 440 CGTTCGTGTGCTATTGCGAGACCTTGGGCTTGGGCGCGGCGGCGGCGGCGGCGGCGGCGG 499
DB 78889 CGCGGACGAGCTGAGAGCGCGTGTGCGACCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 78948
QY 500 GCTGCGCGAGGCGCGCGCGCGCTCGGCTTTGACGAGAGTGTGACACGCTGTTGGCGG 559
DB 78949 CGGCTTCCGCGGCGCGACTGACTTTCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79008
QY 560 CGACCTGACCATGAGAGAGGCGGAGCGAGCTGCGACCGCTCAGCGGCGGCGGCGGCGGCGG 619
DB 79009 CCGCAACGGGATGAACTCGCGCGGTGACCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79068
QY 620 GCGCCACCGGCGACTCGGACGAGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 679
DB 79069 CGCGCGCGGAGAGCGGCGCGCGCGGTGATCGGCTGAGAGCGGCGGCGGCGGCGGCGGCGG 79128
QY 680 CGCTATGCTGAGAAATCTTACCGGAGCTGATCCCTAGTGAAGAGCTGCAAGAGCGG 739
DB 79129 GGAAGT-----CGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79173
QY 740 CCAAGATGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 799
DB 79174 CCGGCGGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79233
QY 800 AAAGGACATGCTGATGCTGCTCATCATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 859
DB 79234 GCTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79293
QY 860 CGACTGCTGCTGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 919
DB 79294 CTACAGAGTTACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79350
QY 920 CGTGAAGCTGGGCAACATCTTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 979
DB 79351 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79410
QY 980 CGAGTGGGCAATTC-----AATGGCGGTGGGCTCGGCGGCGGCGGCGGCGGCGGCGG 1033
DB 79411 CGTACCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79470
QY 1034 CACCGGCGGCTGATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1093
DB 79471 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79530
QY 1094 GCTGCGGCGGCTGAGCTGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1153
DB 79531 GCTGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79590
QY 1154 CAC---CATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1210
DB 79591 CCGCGGCGGTGTGCTCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79650

```

```

QY 1211 CGCGCGCGCGGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1270
DB 79651 CAACCGGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79710
QY 1271 GGGCTTCACCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1330
DB 79711 GGACCTGGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79770
QY 1331 GCTGG 1335
DB 79771 CCGGG 79775

RESULT 4
AAX53491
ID AAX53491 standard; DNA; 114955 BP.
XX AAX53491;
AC
XX
XX
DT 05-JUL-1999 (first entry)
XX
DE Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
XX Antisense oligonucleotide; multiple target; antisense treatment;
XX impaired respiration; inflammation; lung disease;
XX pulmonary vasoconstriction; inflammation; allergic rhinitis;
XX acute asthma; allergy; asthma; impeded respiration;
XX respiratory distress syndrome; pain; cystic fibrosis;
XX pulmonary hypertension; pulmonary vasoconstriction; emphysema;
XX chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
XX colon cancer; breast cancer; lung cancer; pancreatic cancer;
XX hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
XX prostate cancer; ss.
XX
XX Synthetic.
XX
XX PN MO9913886-A1.
XX
XX PD 25-MAR-1999.
XX
XX PF 17-SEP-1998; 98MO-US19419.
XX
XX PR 09-JUN-1998; 98US-0093972.
XX
XX PR 17-SEP-1997; 97US-0059160.
XX
XX PA (UYEC-) UNITV EAST CAROLINA.
XX
XX PI Nyce JW;
XX
XX DR WPI; 1999-229400/19.
XX
XX PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
XX vasoconstriction
XX
XX PS Disclosure; Page 37; 120pp; English.
XX
XX
XX The specification describes antisense oligonucleotides (AAX52869-X55271)
XX directed against at least 2 mRNAs selected from target genes, coding and
XX non-coding regions of RNAs corresponding to target genes, gene
XX initiation codons, genomic flanking regions, intron-exon borders, the
XX 5'-end, the 3'-end and the juxta-section between coding and non-coding
XX regions and all segments of RNAs encoding proteins associated with one
XX or more diseases, conditions or mixtures. The antisense oligonucleotides
XX may be derived from sequences AAX5272-74. These multiple target
XX oligonucleotides (specifically AAX5180-271) can be used for the
XX antisense treatment of diseases and conditions. Typical diseases and
XX conditions are those associated with impaired respiration and
XX inflammation, including lung diseases, pulmonary vasoconstriction,
XX inflammation, allergic rhinitis, acute asthma, allergy, asthma, impeded
XX respiration, respiratory distress syndrome, pain, cystic fibrosis,
XX pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
XX obstructive pulmonary disease (COPD), and cancers such as leukemias,

```

CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
 CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
 CC hepatic metastases, as well as all types of cancers which may metastasize
 CC or have metastasized to the lungs, including breast and prostate cancer.
 XX

SO Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 3.0%; Score 72.6; DB 20; Length 114955;

Best Local Similarity 32.5%; Pred. No. 0.0008;

Matches 411; Conservative 98; Mismatches 742; Indels 14; Gaps 4;

QY 201 CGGCGAGCTCTGCGAGGCGCGGAGTGTGCCCCCGGCTCGCTCGGACGACGAC 260
 DB 104391 CCBGGGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104450
 QY 261 GTGCGTGTAGCCCTTGGCACTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
 DB 104451 NNCGGCGGCG 104510
 QY 321 GCGGCG 380
 DB 104511 SNNNDNNCCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104570
 QY 381 CTGGCCAGCGCGCAAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
 DB 104571 GSNNDNNCCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104630
 QY 441 GTTCG---TGTGCTATTGCGCGAGACCTTGGCGCTGGCGCGCGCGCGCGCGCA 496
 DB 104631 NNNDNNCCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104690
 QY 497 GCAAGCTGGCGGAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 556
 DB 104691 GCBGGCGCGGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104750
 QY 557 CGC---GACCTGACATCATGTGAGGAGGCGAGGAGCTGTGACGCGCTGACCGGCA 613
 DB 104751 CGGNNNDNNCCGCGGCGCGGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCG 104810
 QY 614 CTTGAGAGCG 673
 DB 104811 CCBGGCGGNNNDNNCCGCGGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCG 104870
 QY 674 CTGAGATGCTATGTGAGAAATTTTACCGGAGCTGATCCCTGAGCGAGCTGCAA 733
 DB 104871 SNNNDNNCCGCGGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 104930
 QY 734 GAGCGCGCGATGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
 DB 104931 NCBGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104990
 QY 794 CGCGCGCAAGGACATGTGATGTGTCTATCATCTGCACGCGCAAGAGTGTGAGGC 853
 DB 104991 CGGCGCGGNNNDNNCCGCGGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCG 105049
 QY 854 TGACCGGAGCTGTCTGTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 913
 DB 105050 CGCGCGCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105109
 QY 914 CACCGCGGAGCTGTGCGCAACATCTTCAAGAGAGCGCGCGCGCGCGCGCGCGCG 973
 DB 105110 GGGCGCGCGCGCGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 105169
 QY 974 CGAGGCGGAGTGTGAGCAATCCATGTGAGGTGTGCGCGCGCGCGCGCGCGCG 1033
 DB 105170 BGGCGCGCGCGCGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 105229
 QY 1034 CACCGCGGAGTGTATGTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1093
 DB 105230 GCGCGCGCGCGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105289
 QY 1094 GCTCGCGCGCGCTGAGCGTGTGAGCGGAGTGTGAGCGCGCGCGCGCGCGCGCG 1153

DB 105290 GCGCGCGCGCGCGGCGGNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCG 105349
 QY 1154 CACCATGTGCG 1213
 DB 105350 CGSNNNDNNCCG 105409
 QY 1214 GCG 1267
 DB 105410 CCGCGCGCGGNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105469
 QY 1268 CGCGGCGCTTACACAGAGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCA 1327
 DB 105470 GCGCGGNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105529
 QY 1328 CGGCGTGGCGCAAGCGCGAGAGCTGATCACTAAGATGAGCGCGCGCGCGCG 1387
 DB 105530 GCGTGGCG 105589
 QY 1388 CTTTGTGAGATCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447
 DB 105590 CCG 105649
 QY 1448 CACCG 1452
 DB 105650 CCGCG 105654

RESULT 5.
 AAH13714
 ID AAH13714 standard; cDNA; 1562 BP.
 XX
 AC AAH13714;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:10601.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 PN BP1074617-AZ.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10601; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0166 to AAH1362 and
 CC AAH1363 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 1562 BP; 381 A; 407 C; 449 G; 325 T; 0 other;

XX Query Match 3.0%; Score 71.4; DB 22; Length 1562;

XX Best Local Similarity 52.1%; Pred. No. 0.00075;

XX Matches 159; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

XX 1309 TCGCGTGGCCGTCGCAACGGGCTGGGCAAGGAGCTGATCCAGATGCAGG 1368

XX 1115 TACGCTTGCTGACGCTATGCTTCGAAACATCCAGAAATGATCTGAAGCTTAAGA 1174

XX 1369 CCGGCGAGGCCCAAGTACGATTTGTGAGATCATGGCCCTGCCCGGGCTGTGGGGG 1428

XX 1175 AGGGCAAGTCCATTCCTGCTTTGTGAGAGTCTCTGCTGTGTGAGAGATGCTTAATG 1234

XX 1429 GCGGCGAGGCCCAAGTACGATTTGTGAGATCATGGCCCTGCCCGGGCTGTGGGGG 1488

XX 1235 GCGAGAGGCCCAAGGCTGACGATCCAGAGCATGCGGATTAAGGCCCTGCTGGCGAGATGG 1294

XX 1489 ACAACCTGACGAGAAATCCAGCTGCGCGGAGCCAGCAAGAACCCGTCATCCGAGAC 1548

XX 1295 AAGGCAATTTACGCTGATCCCTGCTGCGGCTCGGAGTCCAGTGCACACGTGCAAGAGC 1354

XX 1549 TGTACGACACGTAACCTCGAGAGCGGCTGGCCAGAGGCCGACGAGCTGTGCAACACC 1608

XX 1355 TGTACGAGAGTGGCTGAGAGGGGATCACTCCGCCAAGGCCGAGAGGTGCTGCATACCA 1414

XX 1609 ACTAC 1613

XX 1415 CGTAC 1419

XX Db

XX RESULT 6

XX AAH27127

XX AAH27127 standard; cDNA; 1577 BP.

XX ID

XX AAH27127;

XX 08-AUG-2001 (first entry)

XX Human NADP hydrogenase subunit 50 cDNA.

XX NADP hydrogenase subunit 50; cancer; haemopathy; HIV infection;

XX immunological disease; inflammation; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 105..1475

XX /product= "NADP hydrogenase subunit 50"

XX MO200130824-A1.

XX 03-MAY-2001.

XX 27-OCT-2000; 2000MO-CN00377.

XX 28-OCT-1999; 99CN-0119893.

XX (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.

XX Mao Y, Xie Y;

XX WPI; 2001-300475/31.

XX P-PSDB; AAB97260.

XX NADP hydrogenase subunit 50 and encoded polynucleotide, applicable in

XX immunological diseases and inflammation

XX Claim 6; Page 25-26; 33pp; Chinese.

XX This invention relates to NADP hydrogenase subunit 50. Included in the

XX invention are nucleotide and amino acid sequences for NADP hydrogenase 50

XX subunit, a vector containing the cDNA sequence, a host cell transformed

XX with the vector, and an antibody directed against the protein. The

XX polypeptide and polynucleotide can be used in the diagnosis and treatment

XX of cancer, haemopathy, HIV infection, immunological diseases and

XX inflammation. The present sequence represents cDNA encoding NADP

XX hydrogenase subunit 50.

XX Sequence 1577 BP; 387 A; 406 C; 457 G; 327 T; 0 other;

XX Query Match 3.0%; Score 71.4; DB 22; Length 1577;

XX Best Local Similarity 52.1%; Pred. No. 0.00075;

XX Matches 159; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

XX 1309 TCGCGTGGCCGTCGCAACGGGCTGGGCAAGGAGCTGATCCAGATGCAGG 1368

XX 1126 TACGCTTGCTGACGCTATGCTTCGAAACATCCAGAAATGATCTGAAGCTTAAGA 1185

XX 1369 CCGGCGAGGCCCAAGTACGATTTGTGAGATCATGGCCCTGCCCGGGCTGTGGGGG 1428

XX 1186 AAGGCAAGTCCATTCCTGCTTTGTGAGAGTCTCTGCTGTGTGAGAGATGCTTAATG 1245

XX 1429 GCGGCGAGGCCCAAGTACGATTTGTGAGATCATGGCCCTGCCCGGGCTGTGGGGG 1488

XX 1246 GCGAGAGGCCCAAGGCTGACGATCCAGAGCATGCGGATTAAGGCCCTGCTGGCGAGATGG 1305

XX 1489 ACAACCTGACGAGAAATCCAGCTGCGCGGAGCCAGCAAGAACCCGTCATCCGAGAC 1548

XX 1306 AAGGCAATTTACGCTGATCCCTGCTGCGGCTCGGAGTCCAGTGCACACGTGCAAGAGC 1365

XX 1549 TGTACGACACGTAACCTCGAGAGCGGCTGGCCAGAGGCCGACGAGCTGTGCAACACC 1608

XX 1366 TGTACGAGAGTGGCTGAGAGGGGATCACTCCGCCAAGGCCGAGAGGTGCTGCATACCA 1425

XX 1609 ACTAC 1613

XX 1426 CGTAC 1430

XX Db

XX RESULT 7

XX AAD17184

XX AAD17184 standard; DNA; 65140 BP.

XX ID

XX AAD17184;

XX 29-NOV-2001 (first entry)

XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX antifungal; antibiotic; nys1; ds.

XX Streptomyces noursei.

XX Key Location/Qualifiers

XX CDS complement (1..1035)

FT /+tag= a
FT /product= "Nysd2 partial protein"
FT /note= "CDS does not include stop codon"
FT complement (1056..2576)
FT /+tag= b
FT /product= "Nysd1 protein"
FT 2806..6906
FT /+tag= c
FT /product= "Nysa protein"
FT 6952..16530
FT /+tag= d
FT /product= "Nysb protein"
FT 16550..49840
FT /+tag= e
FT /product= "Nysc protein"
FT 50260..51015
FT /+tag= f
FT /product= "Nysf protein"
FT 51405..54305
FT /+tag= g
FT /product= "Nysr1 protein"
FT 54329..57190
FT /+tag= h
FT /product= "Nysr2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
FT /+tag= i
FT /product= "Nysr3 protein"
FT 60415..61047
FT /+tag= j
FT /product= "Nysr4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /+tag= k
FT /product= "Nysr5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /+tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /+tag= m
FT /product= "ORF1 protein"
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF-) SINTEF STITTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (EJAE/) FJAEERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX ZOTCHEV SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR,
XX Valle S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides.

PT useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other:
SQ
Query Match 2.9%; Score 70.2; DB 22; Length 65140;
Best Local Similarity 42.5%; Pred. NO. 0.0015;
Matches 573; Conservative 0; Mismatches 756; Indels 18; Gaps 3;
QY 226 AGGTGCCCCCGCGCTCCGCTCGCAGCCAGACCGTCGCTAGCCCTTGACACATTG 285
DB 18456 AGCTCGCCGCGCGCAGCGTCGCGCGGGGTCCTCTCCCTCGAAGACCGCTGACCTCGTCG 18515
QY 286 AGGCGCCGCGCAGCCGCTAGGCAACGTCGCTTGCGCGGCTGCCACCCGCTGGGAGG 345
DB 18516 CCGCCCGCGCGCGCTCAAGAGCCCTGCGCGCGCGCGCGCGATGCTGCGATCCGG 18575
QY 346 CGCCTTTAGTCATGTCACAGCGCGCTCGCGACCTTGCCAAAGCCCAAGACCCCA 405
DB 18576 CCACCGAGAGCAGATCAACCCCACTCACGACGATCTCGATCCCGCCCTCAACG 18635
QY 406 CGCGCAAGCAGCTGCGGTGACGAGTGCCTCGCGCGCTTGTGCTATTGCCAGACC 465
DB 18636 GCGCCACCTCGCTGCTGTCGCGCGCGCGCACGAGGAGCGTCCCGCGATCGGGGCGCGT 18695
QY 466 TGGGCTGCG 525
DB 18696 TCACCGCCCAAGCAGCCCAAGACACCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18755
QY 526 GCTTTGACGAGGTGTTTG-----ACACGCTGTTTGGCGCGCGCGCGCGCGCGCG 579
DB 18756 TCATGACCCGATGCTGCGGGAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18815
QY 580 AGGCGAGCAGCTGCTGACCGCGCTACCGACGACCTGAGAGCGCACCGCGACTCCGACG 639
DB 18816 CGCGCATCCGGGTCTCTCAACCTCACCGGACCGCGCGCGCGCGCGCGCGCGCGCT 18875
QY 640 AGCGGTGCGCATGTTTCACCACTGCTGC---CCGCGTGAATGCGTATGCTGAGGAAT 696
DB 18876 CCGCGACATACGTGGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18935
QY 697 CTTACCGGACCTGATCCCTACGTGAGCAGCTGCAAGAGCCCGCAGATGATGTCGCG 756
DB 18936 CCTCACGACCG 18995
QY 757 CCATGTCAGTCTCTACCTAGCGGAAAGAGGCGATCGCGCAAGACATGTCATCG 816
DB 18996 CCATGCGCCAGGAATCCCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19055
QY 817 TGTTCATATGCTCTGACGCGCGCAAGCACTGGAAGCTGACCGCGCACTGTTCTGTG 876
DB 19056 GCCCGAGAGGCTCTCCCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19115
QY 877 ACGCGACCCCAACCTGCG 936
DB 19116 TCGCTGGGCGCGCGCTTCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19175
QY 937 TCTTCAAGAGCGCGCGCATCAACTGCGCGAGCTGCGCGAGGCGAGTGGACATCCAA 996
DB 19176 CTTTCAGACACAGAGGTTCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 19235
QY 997 TGGCGGTGGGCTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1056
DB 19236 CGGGAATGGGCG 19295
QY 1057 CGGTGCGACGGCGCTATGAGTGTTCAGGGGACGCGCGCGCGCGCGCGCGCGCG 1116

```

Db      19296 CCGGCTACTTTGTTACACGAGCGGCTCTCCGTCGAGACCCCTGCTGCGGACACG 19355
Qy      1117 AGGTGGCGGCATGACGAGCATCAAGAGAACAAATCATCATGATGTCGCCGCCGCG-- 1174
Db      19356 GGGTCCAGGGCCGGGCGCTGCTGCTCCGGGACCCGCTTCTGCAATGAGCTGCTGCGCGCC 19415
Qy      1175 -----GTCMAAGTTTGAAGAGCTGCTGAAGACCGCGCGCGCGCGCGCGAGCGCC 1227
Db      19416 GCGACGAGGCGCGCTGCGACCGCGTGAAGAACTGACCTTGCGCGCCCTGCTGCTGCTG 19475
Qy      1228 CCGCGACGCGACCGCGCGCGCGCGCGCTGCGCTGCGAGCGCGCGCGCGCTTACCAAGAGG 1287
Db      19476 CCGAGCGCGCGCGCGCTGCACTCCAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCC 19535
Qy      1288 ACGGACGCGCGCGCATCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAGCA 1347
Db      19536 GCACTCTCGCATCTTCTCCCGGCTGAGAGACGAGCTTCCGACCTGCTGCTGCGCAACAG 19595
Qy      1348 AGCTGATCAACMAAGATGACAGCGCGCGAGCGCAAGTACGACTTGTGAGATGATGCGCT 1407
Db      19596 CCACCGGCGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 19655
Qy      1408 GCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1467
Db      19656 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19715
Qy      1468 AGAAGCGCGCGCGCGCGCTGTACACCTGACGAGAGTCCACGCGCGCGCGCGCGCGAG 1527
Db      19716 TCGGCTTCCAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19775
Qy      1528 AGAACCGTTCATCCGCGAGCTGTACAGC 1556
Db      19776 AGGTTACGCGCGAAGTGGCGCTGCCGAC 19804

RESULT 8
AAD17186
ID      AAD17186 standard; DNA; 125401 BP.
XX
XX      AAD17186;
XX
DT      29-NOV-2001 (first entry)
XX
DE      Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KM      Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM      antifungal; antibiotic; ds.
XX
XX      Streptomyces noursei.
XX
XX      Key
XX      CDS      Location/Qualifiers
XX      FT      6357..34771
XX      FT      /*tag= a
XX      FT      /product= "Nysi complete protein"
XX      CDS      34792..51099
XX      FT      /*tag= b
XX      FT      /product= "NysJ protein"
XX      FT      51155..57355
XX      FT      /*tag= c
XX      FT      /product= "NysK protein"
XX      FT      57503..58687
XX      FT      /*tag= d
XX      FT      /product= "NysL protein"
XX      FT      complement (58786..58980)
XX      CDS      /*tag= e
XX      FT      /product= "NysM protein"
XX      FT      /note= "CDS does not include start codon"
XX      CDS      complement (59045..60241)
XX      FT      /*tag= f
XX      FT      /product= "NysN protein"
XX      FT      /note= "CDS does not include start codon"
XX      CDS      complement (60238..61296)

```

```

FT      /*tag= g
FT      /product= "NysD2 complete protein"
FT      120628..121308
CDS      /*tag= h
FT      /product= "NysR4 (long) protein"
XX
XX      WO200159126-A2.
XX
XX      16-AUG-2001.
XX
XX      08-FEB-2001; 2001WO-GB00509.
XX
XX      08-FEB-2000; 2000GB-0002840.
XX      10-APR-2000; 2000GB-0008786.
XX      14-APR-2000; 2000GB-0009387.
XX
XX      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX      (SNIF) SINTER STIFTELSEN IND TEK FORSK.
XX      (ALPH-) ALPHARMA AS.
XX      (SINV-) SINVENT AS.
XX      (DZIE/) DZIELEWSKA H.
XX      (ZOTC/) ZOTCHEV S B.
XX      (SEKU/) SEKUROVA O N.
XX      (FJAE/) FJAEVRIK E.
XX      (BRAU/) BRAUTASET T.
XX      (STRO/) STROM A R.
XX
XX      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX      WPI: 2001-557614/62.
XX      P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX      AAE10149, AAE10150.
XX
XX      New nystatin polyketide synthase polynucleotides and polypeptides,
XX      useful as antibiotics and antifungals -
XX
XX      Claim 1: Page 188-254; 266pp; English.
XX
XX      The present invention relates to the cloning and sequencing of the gene
XX      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      involved in the biosynthesis of the macrolide antibiotic nystatin.
XX      The nystatin PKS is useful as antifungal antibiotics. The present
XX      sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX      Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
XX
XX      Query Match      2.9%; Score 70.2; DB 22; Length 125401;
XX      Best Local Similarity 42.5%; Pred. No. 0.0016;
XX      Matches 573; Conservative 0; Mismatches 758; Indels 18; Gaps 3;
XX
Qy      226 AGTGGCGCGCGCGCGCTGCGCTGCGAGCGAGACCGTGGTGAAGCCCTTGCAACTTG 285
Db      78717 AGCTGCGCGCGCGCGAGCGTGGCGGCGCTCTCTCTCGAAGCGCTGACCTCTGCG 78776
Qy      286 AGGCGCGCGCGAGCGCGCTTGAAGCAAGTGGCTTGCAGCGCTGCGCGCGCGAGG 345
Db      78777 CCGCGCGCGCGCGCTCATGAGAGCGCTTCCGCGCGCGCGCGCGAGTCTGCGATCCGC 78836
Qy      346 CGCTTTGAGTATGTCAGACAGCGCGCTGCGCGAGCTTGCCAAAGCCCAAGACGCCCA 405
Db      78837 CCACGAGAGCGAGTACACCCCACTACAGACAGACGCTGATGCGCGCGTCAACG 78896
Qy      406 CCGGCAAGACGCTCTGCGAGAGTGGCTCCGCGCGCTTGTGCTATTTGCGAGACCC 465
Db      78897 GCGCCACTCTCGTGTGCGCGGACCGAGAAACCGTGGCGCGGATGCGGCGCGCT 78956
Qy      466 TGGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCTG 525
Db      78957 TCACCGCGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79016
Qy      526 GCTTTGACGAGGTGTTG-----ACAGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 579

```

```

Db 79017 TCATGAGCCGATGCTGGGAAATTCGGCCCTCCGCGGGGCTGACTTACCAAGAC 79076
QY 580 AGGAGAGCGAGCTGCTGCAACCGCTTCAACGAGCACTGAGAGCCACCGCACTCCGACG 639
Db 79077 CGGCGATCCCGGCTCTCTCCAACTTCAACCGGCAACCGTCCGCGCTGCGCACTGTGCT 79136
QY 640 AGCCGCTCCCATGTTTACCAAGTGTGC---CCGCGTGGATTCGCTATGCTGGAATAAT 696
Db 79137 CCGCCGACTTACTGGGTCCGCAAGTCCGCGAGGCGGTCCCTTCCGCAAGGCGTACCG 79196
QY 697 CTACCCGAGCTGATCTCCCTTACGTAGAGCTGCAAGAGCCCGCAATGATGCTGCGCG 756
Db 79197 CCTTACACGACCCCGCGGTACACAGCTCGTGAATCGGCGCGGAGCGGCTGTCTCCG 79256
QY 757 CCATGCTCAAGTCTTACTAGCGGAAAAGAGGCAATCGCGCAAAAGACATGCTATG 816
Db 79257 CCATGGCGCCAGAAATCCCTGCGGACGCGCCGCGCGCTGCGCTGTCGCAAGAAC 79316
QY 817 TGTTCATCATGCTCTGACGCGCAAGTCCGAGAGCTGACCGCGACTGTTTGTGTGG 876
Db 79317 GCCCGAGAGACTTCTCGCGGTACAGCGGCTGCGCGCGCCAGTCCGCGGCTACCG 79376
QY 877 AGCCGACCCGACCTGCGCGAGCTGGACAGTCAATCAACCACTGAGAGCTGGCAACA 936
Db 79377 TCCGCTGGGCGGCGCTTGTGAGCGGACCGCGCGCGCGCGCGCACTGCCACTAC 79436
QY 937 TCTTCAAGAGCGCGGATCACTGCGCGAGCTGCCGAGGCGAGTGGAGCAATCCAA 996
Db 79437 CTTTCCAGACAGCGGCTTGTGCGCGACCGCGCGCGCGCGCGCGCAAGATCACCGCG 79496
QY 997 TGGGCGTGGGCTGGGCGCGCGCGCTGTGTTCCGACACCGCGCGGTGTATAGAGCGG 1056
Db 79497 CGGACTGGGCGCGCGCGACACCGCTGTGCGCGCAACCGTGAATCTGCGCGAGG 79556
QY 1057 CGTGGCGACGCGCTATGACTGTTTCAAGCGGACGCGCGCGCGCTGAGCTGAGCG 1116
Db 79557 CCGGCTACTTGTTCACAGCGGCTTCTCTCGGACCACTTGTGCTGCGGACGAG 79616
QY 1117 AGGTGCGCGCATGAGCGGATCAAGAGACCAACATCACATGATGTCGCGCGCGCG-- 1174
Db 79617 GGGTCCAGGCGCGGCGCTGCTGCGCGGACCGCGCTTGTGAACTGTCGCGCGCG 79676
QY 1175 -----GTCCAAATTGAGAGAGTGTGAAAGACCGCGCGCGCGCGCGCGAGCGG 1227
Db 79677 GCGACGAGCGCGCTGCGACCGGTGAGGAACTGACCTTGGCGCGCGCTGTGTCTG 79736
QY 1228 CCGCGACAGGACCGCGCGCGCTGAGCTGAGAGCGGCGCGGCTTACCAAGAGAG 1287
Db 79737 CCGAGCGCGCGCGGCTTCAACTCCAGTCCGCGTCCGCGCGCGCGCGCGCGCG 79796
QY 1288 ACGGACAGGCGCGGATCACTGCGCGTGGCGCGGCAAGCGGCTGGGCAAGCGCAAG 1347
Db 79797 GCACCTCGGCACTTCTTCCCGGCTGAGGAGCGGCTTGCACCTGCTGTCCAAACG 79856
QY 1348 AGCTGATCACTMAATAGAGCGCGCGCGGCAAGTCAAGTTCGTTGGAATATATGCGCT 1407
Db 79857 CCAACCGGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 79916
QY 1408 GCGCGCGGCGGTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1467
Db 79917 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79976
QY 1468 AGAAGCGGAGCGCGCGCTGTACAACCTGAGCGAAGAGTCCAGCTGCGCGCGAGCG 1527
Db 79977 TCGGCTTCAAGTACGCGCGCGCGCTTCCAGGCGCTGCGCGCGCGCGCGCGAG 80036
QY 1528 AGAACCGCTCATCGGAGAGCTGTAGAC 1556
Db 80037 AGGTCTAGCGCGAAGTGGCGCTGCCGAC 80065

```

RESULT 9
AAA70438

```

ID AAA70438 standard; DNA; 1836 BP.
AC AAA70438;
XX 28-NOV-2000 (first entry)
DT HIV gp140.mut.mod5F162.delV1/V2 construct coding sequence.
XX
XX HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
XX Human immunodeficiency virus type 1.
OS Synthetic.
XX
XX WO200039302-A2.
XX
XX 06-JUL-2000.
XX
XX 30-DEC-1999; 99WO-US31245.
XX
XX 31-DEC-1998; 98US-0114495.
XX
XX 01-DEC-1999; 99US-0168471.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Barnett S., Zur Megede J., Srivastava I., Lian Y., Hartog K., Liu H.;
PI Greer C., Selby M., Walker C;
XX
XX WPI; 2000-452400/39.
XX
XX Expression cassettes encoding the human immunodeficiency virus (HIV)
PT Gag-containing polypeptide useful for vaccinating against HIV
PT infections and acquired immunodeficiency syndrome (AIDS) -
XX
XX Claim 22; Fig 28; 391pp; English.
XX
XX The present invention relates to synthetic HIV protein expression
XX cassettes. The present sequence is the coding sequence of a synthetic
XX HIV Env expression cassette. In addition, synthetic HIV Gag expression
XX cassettes have been constructed. The Gag protein of HIV is needed for
XX the assembly of virus-like particles. In addition, the Gag protein is
XX involved in many stages of the HIV life cycle, including assembly, virion
XX maturation after particle release and early post-entry steps in viral
XX replication. The expression cassettes of the present invention may be
XX used for the recombinant expression of HIV Gag-polypeptides which may
XX then be used to vaccinate against HIV infection and acquired
XX immunodeficiency syndrome (AIDS).
XX
XX Sequence 1836 BP; 414 A; 623 C; 550 G; 249 T; 0 other;
SQ

```

Query Match 2.9%; Score 69.8; DB 21; Length 1836;
Best Local Similarity 44.3%; Pred. No. 0.0015;
Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

```

QY 462 ACCCTGGGCTGTGGCGCGCGCGCGCACACCCCAACGAGCTGCGAGGCTTCCGCGC 521
Db 544 ACCACGCGATCCGCGCGCTGTGAGACCCAGCTGCTGTGAACCGGACCTGGCGAG 603
QY 522 CTCGCTTTGACGAGGTGTTTGAACAGCTGTGTTGGCGGACCTGACATGAGAGAG 581
Db 604 GAGGGGTGTGTATCCGACGAGGAACTTACCGCAACAGCGCAACATCATCTGTGAG 663
QY 582 GCGAGGAGAGTGTGTGACCGGCTTACCGAGACCTGAGAGCCACCGCGACTCCGAGAG 641
Db 664 CTGAAGAGAGCGGTGAGATCAATGACCCCGCCCAACAAACACCGCAAGAGATC 723
QY 642 CCGCTGCCATGTTTCAACAGCTGCTGCCCGGCTGATGCTATGAGAAATCTTAC 701
Db 724 ACCATGGCGCGCGCGCGCTTCTAGCCTACCGGAGATATTCGCGCATCCGCGAG 783
QY 702 CCGACCTGATTCCTTACGTGAGAGCTGCAAGAGCGCGCAAGTATGCTGCGCGCGAG 761
Db 784 GCCCATGGAATCAAGCGCGCGGAGAGTGAACAAACACCTGTAAGAGATCGTACCAAG 843

```

```

QY 762 GTCAAGTCTTACCTAGCGGAAAAAGAGGCGATGCCCGCAAGACATGTCATGTCTCC 821
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 CTGCAGGCGCCAGTTCTGGCAACAAAGACCATGTGTTCAAGAGAGCGCGCGCACCCC 903
QY 822 ATCATGCGCTGCAAGCGGCAAGGAGTGGAGGCTGACCGGGACCTGTTCTGTGTGACGCG 881
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 GAGATGTGATGCAAGCTTCACTGCGGGCGGCAATTCTTCACTGCAACAGCACCCAG 963
QY 882 GACCCGACCTGCGGCGAGCTGAGCAAGTATCATCACCCTGGAGCTGGGCAACATCTTC 941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 CTGTTCAAGACAGACCTGGAAACAAACACCATCTGCGCCCAACAAACCAAGCGCACATCAC 1023
QY 942 AAGAGCGCGGATCAACTGCGCGAGCTGCGCGAGCGGAGTGGACAATTCATGAGCG 1001
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 CTGCGCTGCGGATCAAGAGATCATCAACCGCTGGCAGAGGTGGGCAAGCGCATGTAC 1083
QY 1002 GTGGGCTGGGGCGCGCGCG---GTGCTGTTGCGACACACCGCGGCTTCATGAGGGGGG 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 GCGCGCGCGCATCGCGCGCAAGATCGCTGCGAGAGCAATCAACCGGCTGTGCTGAC 1143
QY 1059 CTGCGCAGCGCTATGAGCTGTTCAAGGACACGCGCTGCGCGCTGAGAGCTGAGCGAG 1118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 CCGCAGCGGCGGCAAGAGATCAGCAACACCGAGATCTTCCGCGCGCGCGCGCGAG 1203
QY 1119 GTGCGCGGATGACGCGCATCAAGAGACCAACATGACATGTGT---GCCCGCGCGCGG 1175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 ATGCGGCAACATGCGCGCAGAGCTGTCAAGTCAAGTGTGTGAAGATCGAGCCCTG 1263
QY 1176 TCCAGTTGAGAGAGTGTGTAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 GCGGTGCGCCCAACCAAGGCGCAAGCGCGGTGTGTCAGCGCAAGAGCGCGCTGAGC 1323
QY 1236 GGCACCGCGCGCGCGCTGCGCTGGAGAGCGCGCGCGCGCGCTTCCACGAGAGAGCGG 1295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 CTGGGCGCGCATGTTCTTGGGCTTCTGCGGCGCGCGCGCGCGCGCGCGCGCGAG 1383
QY 1296 GCGGCGATCAACTGCGCGTGGCGGTGCGCAAGCGGCTGGGCAAGCGCAAGAGTGTATC 1355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 CTGACCTGACCGTGCA---GGCGCGCGAGCTGTAGCGGAGCTGTGCGAGCAGAGAAC 1440
QY 1356 ACCAAGATGCAAGCGCGCGCGAGCGCAAGTGAAGTGTGTGAGATCATGCGCTG 1408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1441 AACCTGCTGCGCGCATCGAGCGCGCGAGCGACACTGCTGCAAGTGAAGCGGTG 1493

RESULT 10
AAA70434
ID AAA70434 standard; DNA; 1944 BP.
XX
AC AAA70434;
XX
DT 28-NOV-2000 (first entry)
XX
DE HIV gp140.modsFl62.delv2 construct coding sequence.
XX
KM HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
XX
OS Human immunodeficiency virus type 1.
XX
OS Synthetic.
XX
PN WO200039302-A2.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31245.
XX
PR 31-DEC-1998; 98US-0114495.
XX
PR 01-DEC-1999; 99US-0168471.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H,
  Greer C, Selby M, Walker C;

```

```

XX WP1; 2000-452400/39.
DR
XX
PT Expression cassettes encoding the human immunodeficiency virus (HIV)
PT Gag-containing polypeptide useful for vaccinating against HIV
PT infections and acquired immunodeficiency syndrome (AIDS) -
XX
XX Claim 19; Fig 24; 391p; English.
XX
CC The present invention relates to synthetic HIV protein expression
CC cassettes. The present sequence is the coding sequence of a synthetic
CC HIV Env expression cassette. In addition, synthetic HIV Gag expression
CC cassettes have been constructed. The Gag protein of HIV is needed for
CC the assembly of virus-like particles. In addition, the Gag protein is
CC involved in many stages of the HIV life cycle, including assembly, virion
CC maturation after particle release and early post-entry steps in viral
CC replication. The expression cassettes of the present invention may be
CC used for the recombinant expression of HIV Gag-polypeptides which may
CC then be used to vaccinate against HIV infection and acquired
CC immunodeficiency syndrome (AIDS).
XX
SO Sequence 1944 BP; 450 A; 656 C; 577 G; 261 T; 0 other;

Query Match      2.9%; Score 69.8; DB 21; Length 1944;
Best Local Similarity 44.3%; Pred. No. 0.0015;
Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

QY 462 ACCCTGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 ACCCGCGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 711
QY 522 CTGCGCTTGAAGAGGTGTTTGAACAGCTGTTTGGCGCGCGCGCGCGCGCGCGCGCGCG 581
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 GAGGCGTGTGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 771
QY 582 GCGAGAGCTCTGACCGCGCTCAACCGAGCACTGAGAGCGCGCGCGCGCGCGCGCGCGAG 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 CTGAAGAGAGGTGAGATCAACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 831
QY 642 CCGCTGCCATGTTCAACAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 ACCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 891
QY 702 CCGACCTGATCCCTAGCTGAGCAAGCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGAG 761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 GCCCATGCAATACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 951
QY 762 GTCAAGTCTTACCTAGCGGAAAGAGGCGCATCGCGCGCAAGAGACATGTGTGTCTCC 821
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 CTGCAAGGCGCGATGGGCAAGAACATCGTGTTCAGCAGAGCAGCGCGCGCGCGAGCCC 1011
QY 822 ATCATGCGCTGACAGCGCGCAAGAGTGGAGCTGACCGGCACTGTTCTGTGTGACGCG 881
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1012 GAGATGTGATGACAGACTTCAACTCGCGCGCGCGAGTTCTTACTGCAACAGACCCAG 1071
QY 882 GACCCGACCTGCGCGAGCTGAGCAAGTATCATCACCCTGGAGTGGGCAACATCTTC 941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1072 CTGTTCAAGACAGACTTGAACAAACACCATGCGCGCGCGCGCGCGCGCGCGCGAG 1131
QY 942 AAGAGCGCGGATCAACTGCGCGAGCTGCGCGAGCGGAGTGGGCAACATTCATGGCG 1001
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 CTGCGCTGCGGATCAAGCAGATCATCAACGCTGGCAGAGTGGGCAAGCGCATGTAC 1191
QY 1002 GTGGGCTGGGGCGCGCGCG---GTGCTGTTGGCAGCACCGCGGTGTATGAGCGCGG 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1192 GCGCGCGCGCATCGCGCGAGATCGCTGCGACAGCAATCAACCGGCTGTGCTGAC 1251
QY 1059 CTGCGCAGCGCTATGAGCTGTTCAAGGCGAGCGCGCTGCGCGCTGAGCGCGAG 1118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1252 CCGGAGCGCGCAAGAGATCAGCAACACCATCGAGATTTTCCGCGCGCGCGCGAGC 1311
QY 1119 GTGCGCGGATGAGCGCATCAAGAGCAACATCAACCATGT---GCCCGCGCGCGG 1175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

XX XX HIV gp140.mut.modsF162.delV2 construct coding sequence.
DE XX
XX XX HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
KW XX
XX XX Human immunodeficiency virus type 1.
OS XX
XX XX Synthetic.
XX XX WO200039302-A2.
XX XX 06-JUL-2000.
XX XX
XX XX 30-DEC-1999; 99WO-US31245.
XX XX
XX XX 31-DEC-1998; 98US-0114495.
XX XX 01-DEC-1999; 99US-0168471.
XX XX
XX XX (CHIR ) CHIRON CORP.
XX XX
XX XX Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
XX XX Greer C, Selby M, Walker C;
XX XX
XX XX MPI; 2000-452400/39.
XX XX
XX XX Expression cassettes encoding the human immunodeficiency virus (HIV)
XX XX Gag-containing polypeptide useful for vaccinating against HIV
XX XX infections and acquired immunodeficiency syndrome (AIDS) -
XX XX
XX XX Claim 19; Fig 27; 391pp; English.
XX XX
XX XX The present invention relates to synthetic HIV protein expression
XX XX cassettes. The present sequence is the coding sequence of a synthetic
XX XX HIV Env expression cassette. In addition, synthetic HIV Gag expression
XX XX cassettes have been constructed. The Gag protein of HIV is needed for
XX XX the assembly of virus-like particles. In addition, the Gag protein is
XX XX involved in many stages of the HIV life cycle, including assembly, virion
XX XX maturation after particle release and early post-entry steps in viral
XX XX replication. The expression cassettes of the present invention may be
XX XX used for the recombinant expression of HIV Gag-polypeptides which may
XX XX then be used to vaccinate against HIV infection and acquired
XX XX immunodeficiency syndrome (AIDS).
XX XX
XX XX Sequence 1944 BP; 451 A; 655 C; 577 G; 261 T; 0 other;
SQ

```

Query Match 2.9%; Score 69.8; DB 21; Length 1944;
 Best Local Similarity 44.3%; Pred. No. 0.0015;
 Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

```

QY 462 ACCCTGGGCTTGCGCGCGCGCCACACCCCGCAGAGCTGCGCGCGC 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 652 ACCCAGCGCATCCGCCCGCGTGAGACCCAGCTGTGTGAACGGCAGCTGGCGAG 711
QY 522 CTCGGCTTTGACGAGGTGTTTGAACAGCTGTTGGCGCGCAGCTGACCATGAGAG 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 712 GAGGGGCTGTGATCCGACGCGAACTTCCAGCAACCCCAAGACATCATCTGCGAG 771
QY 582 GCGAGCGAGCTGTGACCGGCTCACCGAGCACTGAGGCGCACCGCGACTCGAGGAG 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 772 CTGAGAGAGAGCTGAGATCACTGACCCGCCCAACAACACCCGCAAGAGATC 821
QY 642 CCGCTGCCCATGTTCAACAGCTGCTGCCCGGCTGATGCTATGCTGAGAAATCTTAC 701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 832 ACCATGCGCCCGCGCGCGCTTCTAGGCAACGGGAGATCATGCGCGCATCCGCGAG 891
QY 702 CCGGACTGATCCCTTACGAGAGAGCTGCAAGAGCCCGCAATGATGCTGCGCGCATG 761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 892 GCCCATGCAACATCAGCGCGCGAGAGTGAACCAACCTTGAAGAGATCTGACCAAG 951
QY 762 GTCAAGTCTTACTAGCGGAAAGAGGAGCATGCGCGCAAGAGCATGTCATGCTGTC 821
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 952 CTGAGAGCCCATGTTGCGCAACAGACCATGTGTTCAAGAGAGAGCGCGCGGACCC 1011
QY 822 ATCATGCCCTTGCAAGCGCAAGCATGTGAGGCTGACCGCGCATGTCTGTGTGAGCGCC 861

```

```

DB 1012 GAGATCTGATGATCACAGCTTCAACTGCGCGCGAGTTTCTTACTGCAACAGCACCCAG 1071
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 882 GACCCACCCCTGCGCAGCTGAGACCAAGTCATCACCAGCTGGAGCTGGCAACATCTTC 941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1072 CTGTTCAACAGCACTGGAACAACACCATCTGGCGCCCAACAACACGACCATCTACC 1131
QY 942 AAGAGCGCGGATCAACCTGCGCGAGCTGCCGAGGCGAGTGGGACAATCCATGGAG 1001
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1132 CTGCCCTGCGCATCAAGAGATCATCAACCGCTGCAAGAGTGGGCAAGGCAATGAC 1191
QY 1002 GTGGGCTGCGGCGCGCGC--GTGCTGTTGCGACACCGCGGCTCATGAGCGCGCG 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1192 GCCCGCCCATCGCGCGCGAGATCCGCTGCAGCAGACATCATCGCGCTGTGTGACC 1251
QY 1059 CTGGCAGGCGCTATAGAGTGTTCACAGGACCGCGCTGCGGCTGAGCTGAGCGAG 1118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1252 CCGAGCGCGGCAAGAGATCAAGCAACACCATCGAGATCTTCCGCGCGCGCGCGAGC 1311
QY 1119 GTGGCGGATGAGACGCGATCAAGAGAACCAATCACCATGAT--GCCCGCGCGCGG 1175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1312 ATGCGGACACATGCGGCGAGCGAGCTGTACAAAGTACAAAGTGAAGATCGAGCCCTG 1371
QY 1176 TCCAACTTTAGAGAGCTCTGAAGCACCGCGCGCGCGCGCGCGAGCGCGCGAC 1235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1372 GCGGTGCGCCCGCACCAAGCGCAAGCGCGCGGTGTGACGCGAGAAAGAGCGCGTACC 1431
QY 1236 GCGACCCCGCGCGCGCTGCGCGTGGAGCGCGCGCGCGCTTACCGACGAGAGCGGAGG 1295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1432 CTGGGCTCATGTTCTTGGGCTTCTTGGGCGCGCGCGGACGACATGAGCGCGCGAGC 1491
QY 1296 GCGGCGATCACACTGCGCGCTGCGCGTGGCAACGCGCTGGGCAACCGCAAGATGATC 1355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1492 CTGACCTTACCGTGGA---GCCCGCGCAGCTGTGAGGGGATCTGTGACGACGAGAC 1548
QY 1356 ACCAAGATCAGGCGCGCGGCAAGTACGACTTTGTGAGATCATGCGCTG 1408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1549 AACCTGTCGCGGCGCATCGAGGCGCGACGACCTGCTCAGCTGACCGTGTG 1601

```

RESULT 13
 AAA70433
 ID .AAA70433 standard; DNA; 2025 BP.
 XX
 AC AAA70433;
 XX
 DT 28-NOV-2000 (first entry)
 XX
 DE HIV gp140.modsF162 construct coding sequence.
 XX
 KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.
 XX
 OS Human immunodeficiency virus type 1.
 XX
 OS Synthetic.
 XX
 PN WO200039302-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 30-DEC-1999; 99WO-US31245.
 XX
 PR 31-DEC-1998; 98US-0114495.
 XX
 PR 01-DEC-1999; 99US-0168471.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H;
 XX Greer C, Selby M, Walker C;
 XX
 XX MPI; 2000-452400/39.
 XX
 XX Expression cassettes encoding the human immunodeficiency virus (HIV)
 XX Gag-containing polypeptide useful for vaccinating against HIV
 PT

PT infections and acquired immunodeficiency syndrome (AIDS) -
XX
PS Claim 31, Fig 23, 391pp; English.

CC The present invention relates to synthetic HIV protein expression
CC cassettes. The present sequence is the coding sequence of a synthetic
CC HIV Env expression cassette. In addition, synthetic HIV Gag expression
CC cassettes have been constructed. The gag protein of HIV is needed for
CC the assembly of virus-like particles. In addition, the gag protein is
CC involved in many stages of the HIV life cycle, including assembly, virion
CC maturation after particle release and early post-entry steps in viral
CC replication. The expression cassettes of the present invention may be
CC used for the recombinant expression of HIV Gag-polypeptides which may
CC then be used to vaccinate against HIV infection and acquired
CC immunodeficiency syndrome (AIDS).

5Q Sequence 2025 BP; 478 A; 682 C; 592 G; 273 T; 0 other;

Query Match	2.9%	Score 69.8	DB 21	Length 2025
Best Local Similarity	44.3%	Pred. No. 0.0015		
Matches 422	Conservative	0	Mismatches 522	Indels 9
				Gaps 3

QY	462	ACCCGTGGGCGCTGGGCGCGGGGGCCACCAACCCCAAGACGCTGGCGAGGGCGCTCCGCGC	521
Db	733	ACCCACGGCATCCGCCCGGTGTGACACCCAGCTGCTCTGAAAGGACGCTTGCCGAG	792
QY	522	CTCGGCTTTGACGAGGTGTTTGACACGCTGTTTGGCGCGCACTGACCATCATGAGAG	581
Db	793	GAGGGCGTGTGTATCCGACGGAGAGAACTTCAACCGACAAAGCGCAGACCATCATGTGTGAG	852
QY	582	GGCACGAGCTGTGTGACCGGCTTCAACGAGCACTGGAGGCCACCCGCACTCCGACGAG	641
Db	853	CTGAAGGAGAGCGTGTGAGATCACTGACCCGCCCAACAAACACCCCGAGAGATC	912
QY	642	CCGCTGCCATATTACCAAGCTGTGCGCCCGCGTGGATCGCTATGTGTGAGAAATCTTAC	701
Db	913	ACCATGGCGCCCGCGCGCCTTTACGGCACCGCGGACATCATGTGGCAATCCGCGAG	972
QY	702	CCGACCTGATCCCTTACGTGAGCACTGTCAAGACCCCGAGATGATCTGTGGCGCATG	761
Db	973	GCCCACTGCATCATAGCGGGGAGAAATGGAACAAACACTGTGAAGCAGATCGTCAACAG	1032
QY	762	GTCAAGTCTTACTTACGCGAAAGAGGCGATCGCGCCAAAGACATGTGTATGTGTCC	821
Db	1033	CTGCAGGGCCCAAGTTCGGGCAACAGACCAATCGTGTTCMAAGCAGAGACGGCGGAGACCC	1092
QY	822	ATCATAGCCCTGACCGCGACAGTCGAGGCGTCAACCGCATCTGGTTCTGTGTGAGACCC	881
Db	1093	GAGATCTGTGATGACACAGCTTCAACTTCGCGCGCGAGTTCTTCTACTGCACAGCACCCAG	1152
QY	882	GACCCACACCTTGCGCGACGTGACCAAGTCATCACACCGTGAAGCTGGGCAATCTTC	941
Db	1153	CTGTTCAACAGACCTGGAAACAACACATCGGCCCAACACCAACGAGCACCATCAC	1212
QY	942	AAGGAGCGCGGATCAACCTGCGCAAGCTGCCGAGGGCGAGTGGGACATCTCAATGGGC	1001
Db	1213	CTGCCCTTGCATCAGCAGATCTATCAACCGCTGGCAGGAGTGTGGCAAGGCATGTAC	1272
QY	1002	GTGGGCTCGGGGGCGCGC---GTGTTTTCGGACACACCGCGGTGTATGAGGCGCGC	1058
Db	1273	GCCCCCCCCATCCGGCGCGCAGATCCGCTGTGCACACAAACATCACCGGCTCTGTGTGAC	1332
QY	1059	CTGCGACAGCGCTTATGAGCTTTCACGGGACGCGCTGCGCGCTGAGCTTGAGCGAG	1118
Db	1333	CGGACGGCGGCGAAGAGATCAGCAACACCGAGATCTTCGCGCCCGCGGGCGCGAC	1392
QY	1119	GTGCGGCGATGAGCGGCTCAAGGAGACCAATCACATGCT--GCCCGCGCGGG	1175
Db	1393	ATGCGGACAACTGGCGCGCAGCGATGTACAGTCAAGAGTGTGAMATGAGAGCCCTG	1452
QY	1176	TCCAACTTTGAGAGACTGTGTGAAGCACCGCGCGCGCGCGCGGAGCGCGCGCGAC	1235
Db	1453	GCGGTGGCCCCACCAAGGCGCAGCGCGCGTGTGTGACGCGGAGAACCGCGCGCTGAC	1512

Oy 1236 GGCACCCCCCGGCGCTGAGCTGTGGAGCGCGCGCGGGCTTTCACACACAGAGACGCGAGG 1295
 Db 1513 CTGGGGCCGATGTTCTCTGGGCTTCTGGGGCGCGCGCGACGACCATTTGGGGCCCGCAGC 1572
 Oy 1296 GCGCGCATTCACATCGCCGCTGTGGCCGTGCGCCACGCGCTGGGGCAAGCGCAAGAGCTATC 1355
 Db 1573 CTGACCTGACCGTCA---GCCCCGCACTGCTGACGCGCATCTGTGCACGACGACAGAC 1629
 Oy 1356 ACCAAGATGCAAGCGCGGCGAGGCGCAAGTACGATCTTTGTGAGATGATAGGCTTG 1408
 Db 1630 AACTGTGCGCGGCGCATTCGAGGCGCCACAGACACTGCTGACAGCTGACCGCTGTG 1682

RESULT 14
AAB7043E

AAA70436 standard; DNA; 2025 BP.

AC AAA70436 ;

DT 28-NOV-2000 (first entry)

DE HIV gp140.mut.modS162 construct coding sequence
.....

KW HIV-1; AIDS; Gag; vaccine; expression cassette; Env; ss.

OS Human immunodeficiency virus type 1.

XX

XX.

XX

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

PR 01-DEC-1999; 99US-0168471

PA (CHIR) CHIRON CORP .

Barnett S, Zur Megede J, Srivastava I, Lian Y, Hartog K, Liu H, ...

XX
XX
0000 453400/70

XX

Gag-containing polypeptide useful for vaccinating against HIV

2
.
1
3
7
.
3
9
.
3
2
3
3
.
F
2
L
.
8
5

XX
XX
The association between HIV protein expression

CC cassettes. The present sequence
CC with four generations ago. In

cassettes have been constructed. The gag protein of HIV is needed

CC involved in many stages of the HIV life cycle, including assembly, viral

CC replication. The expression cassettes of the present invention may be used for the overexpression of HIV Gag-polymers which may

CC then be used to vaccinate against HIV infection and acquired

[illegible]

08-09-2025 DB 21: 10:00

Best Local Similarity 44.3%; Pred. No. 0.0015;

A62 **XXXXXXXXXXXXXXXXXXXXX**

713 ACCGATCCCTGAGGACGCTGAGCGAG 7

Sequence 2025 BP; 479 A; 681 C; 592 G; 273 T; 0 other;

Query Match	2.9%	Score 69.8	DB 21,	Length 2025,
Best Local Similarity	44.3%	Pred. No. 0.0015,		
Matches 422; Conservative	0;	Mismatches 522;	Indels 9;	Gaps 3;

[illegible]

```

OY 522 CTCGGCTTTGACGAGGTGTTTGAACAGCTGTTTGGCGCCGACCTGAACCATCATGAGAG 581
DB 793 GAGGGCGTGTGATCCGACGAGAACTTACCGAACGCCAAGACCATCATCTGTGAC 852
OY 582 GGGAGGAGCTGTGTGACCGGCTTACCGAGCACTGTGAGGCCCAACCCGACCTCCGACGAG 641
DB 853 CTGAAGAGAGCGGTGAGATCAACTGCAACCCGCCCAACAAACCCGCAAGAGCATTC 912
OY 642 CCGCTCCCATGTTTCAACAGCTGTGCCCCGGGTGTGATGCTATGTGAGAAATCTTAC 701
DB 913 ACCATGGGCCCCGGCGCGCTTCTAGCCACCGGAGCAATCATCGGCGACATCCGCCAG 972
OY 702 CCGGACCTGATCCCTTACCTGACAGCTGCAAGAGCCCCCAATGATGTGCGGCGCATG 761
DB 973 GCCCACTGCAACATCAAGCGCGGAGAAAGTGAACAAACCTTGAAGAGATCTGTGACCAAG 1032
OY 762 GTCAACTCTTACTAGCGGAAAGAGGATGGCCCAAGAGCATGTGATGTGTTC 821
DB 1033 CTGCAAGGCCGATTTGGCAACAAAGACCATGTGTTCAGGACGAGCGCGCGGACCC 1092
OY 822 ATCATGCCCTGACGCGCAAGCAGTGGAGCTGACCGGACATGTTCTGTGTGACGCC 881
DB 1093 GAGATGTATGACAGCTTCAACTGCGGCGGAGATTCTTCTACTGTCAACAGACCCAG 1152
OY 882 GACCCCACTGCGGACGCTGACCAACGTCATCAACACCGTGGAGCTGGCAATCTTC 941
DB 1153 CTGTTCACACAGCACTGGAACCAACACCATGTGGCCCAACAAACCAACGACCATTCAC 1212
OY 942 AAGGAGCGGGCATCAACTGTGGCGGAGCTGCCGAGGGCGAGTGGGCAATCCAAATGGGC 1001
DB 1213 CTGGCTGCGCATCAAGAGCATCAACCGTGGCAGAGGTGGGCAAGGCGCATGTAC 1272
OY 1002 GTGGGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
DB 1273 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1332
OY 1059 CTGGCGACGCGCTTATGAGCTGTTCACGCGGACCGCGCGCGCGCGCGCGCGCGCG 1118
DB 1333 CCGGACGCGGCGGCAAGAGATCAAGCAACACGAGATCTTCGCGCGCGCGCGCGCGG 1392
OY 1119 GTGGCGGCGATGAGACGCGCATCAAGAGAGCAATCAATCAATGTGT---GCGCGCGCGCG 1175
DB 1393 ATGCGGACAACTGGGCGAGCGAGCTGTACAAATGATGATGATGATGATGATGATGAT 1452
OY 1176 TCCAAATTTGAGAGAGCTGTGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1235
DB 1453 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1512
OY 1236 GCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
DB 1513 CTGGGCGCGCATGTTCTGTGGGCTTCTGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 1572
OY 1296 GCGGCGATCACTGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
DB 1573 CTGACCTGACCGTGA---GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1629
OY 1356 ACCAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
DB 1630 AACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1682

```

RESULT 15

AAAS1944 standard; DNA; 2310 BP.

AAAS1944;

31-OCT-2000 (first entry)

Modified HIV-1 Env gp160, Val120-A1a204.

HIV-1; vaccine; anti-HIV; immunogenic; modified; envelope; Env;
 CD4 binding region; V1/V2 loop; bridging sheet; ss.

```

XX Human immunodeficiency virus type 1 isolate SF162.
OS Synthetic.
PN MO200039303-A2.
PD 06-JUL-2000.
XX 30-DEC-1999; 99MO-US31272.
XX 31-DEC-1998; 98US-0114495.
XX 29-SEP-1999; 99US-0156670.
PA (CHIR ) CHIRON CORP.
PI Barnett S, Hartog K, Martin E,
DR WPI; 2000-465745/40.
XX
XX Novel modified HIV Env polypeptides useful as immunizing agents and for
PT preparing a vaccine to elicit an immune response against a broad range
PT of HIV subtypes
XX
XX Claim 14; Page 118; 139pp; English.
PS
XX
XX Novel immunogenic modified human immunodeficiency virus (HIV) envelope
CC (Env) polypeptides having an amino acid deleted or replaced in the region
CC corresponding to residues 420-436 or 119-123 and 199-210 relative to
CC isolate HXB-2 are disclosed. The modified Env polypeptide is based on HIV
CC strain SF162, with numbering relative to isolate HXB-2. The Env
CC polypeptides are modified so as to expose at least part of the CD4
CC binding region. The modified HIV Env polypeptides, coding polynucleotides
CC and constructs, further comprising an adjuvant, are used for inducing an
CC immune response in an individual. The method involves administering a
CC first composition comprising a polynucleotide encoding the Env
CC polypeptide in a priming step and administering a second composition
CC comprising a modified Env polypeptide as a booster in an amount
CC sufficient to induce an immune response in the individual. The first
CC and/or second composition further comprises an adjuvant (claimed). The
CC intracellularly produced Env polypeptides can be used for a number of
CC diagnostic and therapeutic purposes to determine the presence of reactive
CC antibodies and/or Env proteins in a biological sample to aid in the
CC diagnosis of HIV infection or disease status or as measure of response to
CC immunization.
XX
XX Sequence 2310 BP; 477 A; 805 C; 707 G; 321 T; 0 other;

```

Query Match 2.9%; Score 69.8; DB 21; Length 2310;

Best Local Similarity 44.3%; Pred. No. 0.0015;
 Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

```

OY 462 ACCCTGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
DB 496 ACCCAAGGATCCGCCCGGTGTGAGACCGACCTGTCTTGAACGCGCGCGCGCGCG 555
OY 522 CTCGGCTTTGACGAGGTGTTTGAACAGCTGTTTGGCGCCGACCTGAACCATCATGAGAG 581
DB 556 GAGGGCGTGTGATCCGACGAGAACTTACCGAGCAAGGCCAAGCATCATCTGTGACG 615
OY 582 CCGCTCCCATGTTTCAACAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
DB 616 CTGAAGAGAGGCGTGAAGATCAACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
OY 642 CCGCTCCCATGTTTCAACAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
DB 676 ACCATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
OY 702 CCGGACCTGATCCCTTACCTGAGAGCTGCAAGAGCCCCCAATGATGTGCGGCGCATG 761
DB 736 GCCCACTGCAACATCAAGCGCGGAGAGTGAACAAACCTTGAAGAGATCTGTGACCAAG 795
OY 762 GTCAACTCTTACTAGCGGAAAGAGGATGCGCGCAAGAGCATGTGTGTGTTC 821

```

```
Db 796 CTGCAGGCCCGATTGCGCAACAAGACCATCGTGTTCAGCAGACAGCGGCGGACCCC 855
QY 822 ATCATGCCCTGCAGCGCAAGCAGTCGAGGCTGACCGCGACTGTGTTCTGTGTGAGCGCC 881
Db 856 GAGATCGTATGCAACAGCTTCAACTGCGGGGGAGTTCTTCTACTGCAACAGCACCAG 915
QY 882 GACCCCACTGCGCGCAGCTGAGCAACAGTCATCAACCGTGGAGCTGGCAATCTTC 941
Db 916 CTGTTCAACAGCACCTGGAAACAACCATCGGCCCAACAACAACCAACGACCATCACC 975
QY 942 AAGAGCGCGGCGCATCAACCTGCGCGAGCTGCCAGGGGAGTGGGACAATCCAATGGGC 1001
Db 976 CTGCCCCCTGCGCATCAAGCAGATCATCAACCGCTGCGAGAGTGGGCAAGGCCATGTAC 1035
QY 1002 GTGGGCTCGGGCGCGCGC---GTGCTGTTGCGCACCAACCGCGGTGTATGAGCGCGCG 1058
Db 1036 GCGCGCGCGCATCGCGCGCGCGAGATCCGCTGCAAGCAGCAACATCAACCGGCTGTGTGACC 1095
QY 1059 CTGCGGACGGGCTATGAGCTGTTCAAGGCGACGCGCGCTGCGGCTGAGCCTGAGGAG 1118
Db 1096 CGCGACGCGCGCAAGGAGATCAGCAACACCAACGAGATCTTCCGCGCGCGCGCGGAGC 1155
QY 1119 GTGCGGCGCATGAGACGCGCATCAAGAGACCAACATCATGATGAT---GCCGCGCGCGGG 1175
Db 1156 ATGCGGACAACTGGCGCAGCAGCTGTACAGTACAGAGTGTGAAGATCGAGCCCTG 1215
QY 1176 TCCAAAGTTGAGAGCTGTGAAGCACCGCGCGCGCGCGCGCGCGAGCGCGCGCAC 1235
Db 1216 GCGGTGCGCGCGCACCAAGGCGCAAGCGCGCGGTGTGCAAGCGGGAAGCGCGCGTGTACC 1275
QY 1236 GGCACCGCGCGCGCGCTGCGCTGGAGCGCGCGCGCGGCTTCAACAGCAGAGAGCGCAGG 1295
Db 1276 CTGGGCGCGCATGTCTCTGCGGCTTCTCTGCGCGCGCGCGCAGCACCATGGCGCGCGCAGC 1335
QY 1296 GCGGCGATCACTGCGCGGTGGCGCGTGGCCCAACGCGCTGGGCAACGCCAAGAGCTGATC 1355
Db 1336 CTGACCTGTGACCGTGA---GGCGCGCAGCTGTGAGCGGCATGTGTGACAGCAGAAC 1392
QY 1356 ACCAAGATGAGCGCGCGAGGCGCCAGTACGACTTTGTGAGATCATGGCCTG 1408
Db 1393 AACCTGCTGGCGCGCATGAGGCCCAAGCAGACACTGTGACGACGTGTG 1445
```

Search completed: June 8, 2003, 10:17:54
Job time : 857 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2003, 10:01:16 : Search time 139 Seconds
(without alignments)
5292.932 Million cell updates/sec

Title: Af012098

Perfect score: 2399
Sequence: 1 ATCTTACATGACACACAAA.....TGTAGCAGCAGCTTCTGAG 2399

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65.4	2.7	44377	2	US-08-804-227C-7
2	65.4	2.7	44377	2	US-08-804-198-1
3	64.6	2.7	50937	4	US-09-428-517-1
4	63.4	2.6	1182	4	US-09-385-028-19
5	63.4	2.6	11604	4	US-09-385-028-13
6	63.4	2.6	15079	4	US-09-385-028-11
7	62.6	2.6	1217	4	US-09-372-422A-11
8	61.8	2.6	735	3	US-09-003-287-7
9	61.8	2.6	8438	1	US-07-945-283-1
10	61.4	2.6	5392	2	US-08-403-852D-1
11	61.4	2.6	5392	4	US-08-510-646B-1
12	61.4	2.6	5392	4	US-09-231-818-1
13	61	2.5	2713	3	US-08-804-439A-13
14	61	2.5	2713	3	US-08-720-229-13
15	60.4	2.5	4403765	4	US-09-103-840A-2
16	60.4	2.5	4411529	4	US-09-103-840A-1
17	60.2	2.5	2721	6	5215881-2
18	60	2.5	1288	1	US-08-440-856A-9
19	59.4	2.5	1605	4	US-09-124-541-3
20	59.4	2.5	3472	6	5244792-2
21	59.4	2.5	6733	4	US-09-124-541-2
22	59.4	2.5	13842	4	US-09-105-537-30
23	59.4	2.5	36778	4	US-09-105-537-5
24	59.4	2.5	38506	4	US-09-320-878-19
25	58.8	2.5	15872	4	US-09-105-537-1
26	58.4	2.4	4403765	4	US-09-103-840A-2
27	57.8	2.4	1206	4	US-09-372-422A-13

28	57.8	2.4	11220	4	US-09-105-537-32	Sequence 32, Appl
29	57.4	2.4	1755	1	US-08-068-395A-2	Sequence 2, Appl
30	57.4	2.4	1755	1	US-08-464-365-2	Sequence 2, Appl
31	57	2.4	999	2	US-08-761-277A-46	Sequence 46, Appl
32	56.8	2.4	2277	1	US-08-676-967-5	Sequence 5, Appl
33	56.8	2.4	2277	1	US-08-676-974-5	Sequence 5, Appl
34	56.8	2.4	2277	2	US-09-098-487-5	Sequence 5, Appl
35	56.6	2.4	1288	1	US-08-440-856A-9	Sequence 9, Appl
36	56.2	2.3	4257	2	US-08-690-473-1	Sequence 1, Appl
37	56.2	2.3	4257	4	US-09-259-821A-1	Sequence 1, Appl
38	56.2	2.3	4257	4	US-08-843-659-1	Sequence 1, Appl
39	56.2	2.3	12001	1	US-08-458-568A-11	Sequence 11, Appl
40	56	2.3	49377	1	US-08-764-233A-1	Sequence 1, Appl
41	55.8	2.3	987	4	US-09-159-106-12	Sequence 12, Appl
42	55.8	2.3	1516	4	US-09-159-106-10	Sequence 10, Appl
43	55.2	2.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
44	55	2.3	3468	1	US-07-951-715A-2	Sequence 2, Appl
45	55	2.3	3468	2	US-08-459-448A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kohnstoss, Stuart A.
; APPLICANT: Ruston, Paul R., Jr.
; APPLICANT: Suteck, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: .350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:


```

NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match
Best Local Similarity 2.7%; Score 65.4; DB 2; Length 44377;
Matches 502; Conservative 0; Mismatches 641; Indels 12; Gaps 4;

QY 475 CCGCGGCGCCACACACCCCAAGCAGCTGAGCCGAGGGGCTCCGCGCCCTGCGCTTTGAGC 534
DB 29120 CCGAGGCGCCCGGACCGCTGCTGAGAGTGGCCCGGCGCTCACCGGTGTCGCTGGGG 29179
QY 535 AGGTGTTTGAACAGCTGTTTGGCGCGACCTGACATCATGAGAGGCGACGACTGC 594
DB 29180 ACCGCGTGTGCGGCTGTGGGAGCGCGGCTGGGGCCGCTGTGCTGGCCGACCGCG 29239
QY 595 TCGACCGCTTCAACCGAGACCTGAGAGGCCACCCGCACTCCAGCGAGCGGCTGCCATGT 654
DB 29240 TCGTGGCCCCCGTCCGAGACGGCTGTCTTACCGCCAGGCGGCTGCTCCGCGGTGT 29299
QY 655 TCACACGCTGCTGCCCCCGCTGAGTGTATGCTGAGAAATCTTACCGACCTGATCC 714
DB 29300 TCTCTAGCCCTACTACTAGCTGTGTACCTGCGCGCTCAGAGCGGGGGAGCGGGTGC 29359
QY 715 CCTACGTGAGCAGCTGCAAGAGCCCCCAATGATGCTGGCGGCGATGTCAGTCTTACC 774
DB 29360 TCGTGACACCGCGCGCGGGGGGCGTGGCATGCGCGGTGAGATGCGCCGCGACCTCG 29419
QY 775 TAGCGAAAAAGAGGGCATGCGCCCAAGACAT---GGTCATGTGTCTCATATGCTCT 831
DB 29420 GGGCGGAGGTGCTGGCCACCGCAGGCCCGGCAAGTGGAGCGCCCTGCGCGCATGGGA 29479
QY 832 GCACGCGCAAGCAGTGGAGGCTGACCGGCACTGTTCTGTGTGAGCGCGCAACCCACCC 891
DB 29480 TCACGACACACACACTCTGCTCTCTCCGACCTCGACCTTTCGACCGCTTACCGGAG 29539
QY 892 TCGCGCAGCTGAGCAACGTCATACCAACCGTGA---GCTGGGCAACATCTTCAAGAG 948
DB 29540 CGGACCGGACGCTCCCGCGGAGCGTGTCTGAACTCGCTCACCAAGAGTGTGTGGAGC 29599
QY 949 GCGGCACTCAACTGCGCGAGCTGCCGAGGGGCAAGTGGCAATCCAAATGGGCTGGGCT 1008
DB 29600 CTTCTCTGGGCTGCTCCGTCGCGGCGCGGCTTCTGTGAGCTGGGCAACGACGTCG 29659
QY 1009 CGGCGCCGAGCGTGTGTTTGGGACCAACCGGCGGTGTCATGAGGCGCGCTGCGACGG 1068
DB 29660 GGGACCCCGAGCGGATGCGCGGAAACCCCGGGGTGCGCTACCGGCTTTCGACCTCA 29719
QY 1069 CCTATAGCTGTTCAAGGCAACCGCTGCGCGCTGAGCCTGAGCGAGTGGCGGCA 1128
DB 29720 ACGAGCGCGGAGCCGAGCACTGCGCGGCTGTGGGGGAACTGATGGAATCTTTGGCG 29779
QY 1129 TGGAGCGCATCAAGAGACCAACATCACTGTTGCCGCGCGGCTTCAAGTTTGAAG 1188
DB 29780 CCGGCTGTGTCACCGCTGCGCTGCTGTCAACCCACGACGTCGCGCGCGCG---GACG 29836
QY 1189 AGCTGTGAAGCAACCGCGCGCGCGCGGCGAGGCGCGGCAACCGGACCCCGCGCG 1248
DB 29837 CCTTGGCACATCAAGCTAGGCGCGCGGCAACCGGAAAGCTGTCTTGAACATGCGCGCG 29896
QY 1249 CGCTGCTTGGGACCGGCGCGCGGCTTCAACGAGGAGCGGCGAGGGGCGGATCAAC 1308
DB 29897 CTTGGGACCGGTACCGGACGGTCTGTGTCACTGGTGGCACTGGGCGCGCTGGGAGCGCA 29956
QY 1309 TCGCGTGGCGCTGGGCGACGCGGCTGGGCAACCGCAAGAGTGTATCAACCAATGACG 1368
DB 29957 TCGCCCGCACTGGGAGCGGCGGCGGCTGCGCGGCTGTGATGCGCGCGCGCGG 30016
QY 1369 CCGGCGAGGCGCAAGTCACTTTGTGAGATCAATGGCTGCGCGCG---GGGCTGTGG 1425
DB 30017 GCTCGGACCGGCGAGGGGCGCGGAGGCTGTGCGCGCACTTGGCGCGCTTGGGCGCGTGG 30076
QY 1426 GCGGCGCGCGCGCGCGCTCCACGCAAGGCACTACGCAAGCGGCGAGGCGCGCG 1485

```

```

DB 30077 CCACGCTGCTCCCTGCGACGCTCTCCGAGCGGAGCCCTCCGCGACTGCTGCGGCA 30136
QY 1486 TGNACACCTTGAGCAGGAGTCCACGCTGCGCGCGGACCGGACGAGAACCCGTCATCGCG 1545
DB 30137 TACCGGCGATACCCGCTGACGCGGTGTGTACAGACACCGCGCTCTTCAGACGCGCG 30196
QY 1546 AGCTGTACGACACGTAATCTGAGAGCCGCTGGGCGCAACAGGCGGACGAGCTGTGACA 1605
DB 30197 TCGTCCGCGGCTCAACCCCGAGCGGAGTGGCGGCTGTGCGGCGCAAGTGTGAGGCGG 30256
QY 1606 CCCACTACGTGGCGG 1620
DB 30257 CCGTCACCTGGAGC 30271

RESULT 3
US-09-428-517-1/c
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bellach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; EARLIER FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER FILING DATE: 60/106,100
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match
Best Local Similarity 2.7%; Score 64.6; DB 4; Length 50937;
Matches 376; Conservative 0; Mismatches 419; Indels 30; Gaps 3;

QY 857 CCGCGACTGTTCTGTGTGAGCGCGGACCCACCTGCGCGAGCTGGACCACTCATCAC 916
DB 3774 CCGCCCTCGGAGGAGTGTTCGCCGCGGTGGGGGCGGTGTCCTTACGACGCGACCAT 3715
QY 917 CACCTGAGAGCTGGGCAATCTTCAAGAGCGCGGCACTCAACCTGGCGGAGCTGCCGA 976
DB 3714 CACCTGCGGCTACCGGAGCTACCGCATCGCGCGCGCGCGCGGCAAGCGCTTA 3555
QY 977 GGGCGAGTGGACCAATCAATGAGCGCTGGGCTTGGGCGCGCGGCTGTTCGGACAC 1036
DB 3654 CCGTGAACCTAGCGGCGGAGAGCGCTTGGCGCGTGGCTTACCGGACGACTCAG 3595
QY 1037 CCGCGGTGTATGAGGCGGCGCTGCGACCGGCTATGAGCTGTTCAGGGGACGCGCT 1096
DB 3594 CAGCGCCCATATGCGGCTGGCGGCGGAGACGCAAGACTGCGCATCACGCGCGCGCT 3535
QY 1097 GCGCGCGCTGAGCGTGAAGAGTGGCGGCACTGACCGGATCAAGAGACCAACATCAC 1156
DB 3534 GGGGAAGTGGCCCATATCACTTGAAGTGAAGTGTGACTCCGACAGCGCGCTCAGAGAGC 3475
QY 1157 CATGTGCGCGCGCGCGGCTCCAAATTTGAGAGCTGTGAAGCAACGCGCGCGCGCG 1216
DB 3474 GATGACCGGG---CGGTGTCAATGCGCGGAGAGGAGAGACCGCGGTCTCTTCGCTGA 3418
QY 1217 CCGCGAGCGCGCGCGGACGCGACCCCGCGCGCTGTGACTGTGGAGCGGCGCGCGCTT 1276

```



```

Db      3417 CCAAGCGGCTCGGAGACGGGACCGACGTGCTGCTCCGCGCCGCCAGAGACA 3358
QY      1277 CACACGAGGAGACGGAGGGCGGATCACTACGTGCGCGTGGCCGCA----- 1327
Db      3357 CCGATCAGGCGCGTGGCGCGGTCAACCCAGGACCGTGTGTCTCAACACGGGCTC 3298
QY      1328 -----CGGCTGGGCAACGCCAAGAGTATACCAAGATGACGGCGG 1372
Db      3297 CTCCTCACCATTCCCTGTGTGAGAGAACCGCGCGGTCTCACCATTGTGATCCCGG 3238
QY      1373 CGAGCCAGATGACTTGTG---TGAGATCATGGCTGCGCGCGGCTGTGTGGCGG 1429
Db      3237 CAGGCGCGGCGCGAGGCGCACCGCGCTGTGTTGTGGAGCGCGACCCAGGAGCGG 3178
QY      1430 CGGCGCGACGCGCGCTGCACCGACCAAGGCCATCAGCGACAGCGGCGCGGCTGTA 1489
Db      3177 GGTACCCAGACTTCCCGCGCGACGAGGGGAGAGCGCGTTCGCGGGAGCCGAGGCG 3118
QY      1490 CAACCTGAGAGAAAGTCCACGCTGCGCGCGGACCGACAGAAACCGTTCATCCGCAAGT 1549
Db      3117 CTACCGCGGAGTGTGATGACCACTGATCTCCTCGAGGCGATCTACTCCGCTACCGCTG 3058
QY      1550 GTACGACAGTACTCTCGAGAGCGCGTGGGCAAAAGCGCAGAGCTGTGACACCA 1609
Db      3057 GTACGACAGAGAGGCGCTCCAGCGCGCTGTCTCTTCTGCGCACGGTTGTGCTACCTC 2998
QY      1610 CTACGTGCGCGCGCGCTGAGAGAAAGACGAGAAAGTGAAG 1654
Db      2997 CTTCGACTACCGGAGCTGAAAGTGAAGGAGCGGAGCGGCGAGCGGCGG 2953

RESULT 4
US-09-385-028-19
; Sequence 19, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwame A. Aido
; APPLICANT: Ashish S. Parakh
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jennifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1182 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-385-028-19

Query Match      2.6% Score 63.4; DB 4; Length 1182;
Best Local Similarity 43.5%; Pred. No. 0.0013;
Matches 475; Conservative 0; Mismatches 586; Indels 30; Gaps 3;

QY      405 ACCGCCAAGCAGCTGTGCTGAGAGTGTGCTCCGCGCTTCTGTGCTATTGCGAGACC 464
Db      22  ACCGCCCGGAGATTGTGTGACACAGGCGCGGTGGGTCTGGCCGAGACGGCGCGAC 81
QY      465 CTGGGCTGGCGCGCGCGCGCACACACCCCAAGACGTGGCCGAGAGGCTTCCGCGCTC 524
Db      82  GACTTACCGCTCTGCGCTCCACCGCGCGGACCGGTGAGCGCGCTTTCACCGCTCC 141
QY      525 GCGTTGACAGAGTGTGACACGCTGTTTGGCGCGACCTGACCATGTGAGAGAGGCG 584
Db      142  CGCTTGGCGGCGCGAGCGTCTGTGTGCGGAGGCGGTGCGACCGGCGAGGCGCGC 201
QY      585 ACGAGCTGTGCTGACCGCTTCAACGAGCACTGAGGCCCAACCGCATCTCGACGAGCGG 644
Db      202  GGTGTGTGTGTCTGTGCGCGCGCAACGGAAATGTGGGACCGCGCTGGAGGCGGAGAAC 261
QY      645 CTGCCCATTTTACCAAGCTGTGCTGCGCGCGCTGATGCTATGTGAGAAATCTTACCGG 704
Db      262  GCGCGGAGAGTGTGAGAGCGCGTGTGCGCGCGCTGCGGCTCGGAGGCGAGATGCTG 321
QY      705 GACCTGATCCCTAGTGTGAGAGCTGCAAGAGCGCCCGACATGATGTGCGCGCATGTG 764
Db      322  ATGCGCTTCAACCGGCTGTATGCGCGCGCATACCGATGAGAGATCTCGGAGCACCTC 381
QY      765 AAGTC-----CTACCTAGCGGAAAAAGAGGCGATGCGCCAAAGAGATGTGATGTG 818
Db      382  AAGAGCTGTGAGTGTGCGCGCGCGCGGAGGCGGCTTTCGACCGCGCGCGCATCATG 441
QY      819 TCCATATGCTGTGACGCGCGCAAGCAAGTGTGAGGCTGACCGGACTGTGTGTGTGAG 878
Db      442  ACGACCGAGACCGCGCGCGCAAGAGAGTGTGCGGTGATGAGTGTGCGGCGGCGACCTC 501
QY      879 GCGGACCCCACTGTGCGCGAGTGTGACGAGCATATCACCGGTGAGTGTGCGCAATC 938
Db      502  ATGCGCAAGGCGGTGTGCGCATGTGAGCGCGGACATGCGAC--GCTGTGACTTTCTT 558
QY      939 TTCAAGAGCGCGCGCATCAACTGTGCGCGAGCTGTGCCGAGGCGAGTGGGACAAATCCAATG 998
Db      559  GCGAGGAGCGCGCGCTGTGACCGCGCGCGAGCAGACCGCTCTTCCGCGGTCATGAG 618
QY      999 GCGGTGGCTGTGCGCGCGCGCGGTGTGTGCGCACCAACCGCGGTGTATGAGAGGCGG 1058
Db      619  CGGACCTTCAACCGCGGTGTGAGCATGTGACACCGAACCTTCAACGACGACGCGGCTG 678
QY      1059 CTGCGCAGC-----GCTATGAGCTGTTCACGGGCGAGCGCGCTG 1097
Db      679  TTGCGCAAGGCGCTGTGCGCGCGGAGGTGTGACGCGCGGAGTGTGAGAGGCGCTG 738
QY      1098 CCGCGCTGAGCTGTGAGGAGTGTGCGCGCATGTGAGCGGATCAAGAGACCAATCATACC 1157
Db      739  GCGCGCTGTGCGCTGTGCAAGGACATGTGAGAGGAGCGGAGGCGCGCGCAAGCTGATC 798
QY      1158 ATGTGCGCGCGCGCGCGGTTCAAAGTTTGTGAGAGCTGTCTAAGACCGCGCGCGCGG 1217
Db      799  GAGGTCAAGTTCACCGCGCGCGCGAGAGCGGACGCGGCAAGGCGGTGTGCAAGACCTG 858
QY      1218 GCGGAGCGCGCGCGCGACGAGCAACCGCGGCGCTGAGTGTGAGCGCGCGCGCGCTTCT 1277
Db      859  GTAACTCTCCCTGTTGTGTAAGACCGCGGTGTGACGCTGTGACCGCCCACTGTGGGCG 918
QY      1278 ACCAGGAGAGACGAGGCGCGCATCATGCGCGGTGTGCGCAACGGGCTGTGGGCG 1337
Db      919  GCCATGGCATGTGCGCAAGTGTCTGAGAGACACCGATGACACGAGGCGGCTGAGCATC 978

```

QY 1338 AACGCCAAGAGTGTATCAACCAAGATGACGGCCGGCCGAAGTACGACTTTGTGGAG 1397
DB 979 CCCTTCGGAGAGGTGAGGTCTATCCGCCGAAGGCCCGGAGCCGACGAGCCGACGCG 1038
QY 1398 ATCATGAGCTGCTGCCCCGGGGCTGTGGGCGGGGGCGGAGCCCGCTTCCACCGACAG 1457
DB 1039 CTGGGGGCGCGCTCCGGAGCATCTGCGGGCGGACGAGGTGTCTATCGGAGTGAACCTC 1098
QY 1458 GCCATCAGCA 1468
DB 1099 GCCATCGCGA 1109

RESULT 5
US-09-385-028-13
; Sequence 13, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A. Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jennifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-13

Query Match 2.6%; Score 63.4; DB 4; Length 11604;
Best Local Similarity 43.5%; Pred. No. 0.00022;
Matches 475; Conservative 0; Mismatches 586; Indels 30; Gaps 3;

QY 405 ACGCGAAGCACTGTGCTGACGAGTGTCCGCGCTTGTGCTATTGCGGAGAC 464
DB 5884 ACGCCCGGGGATTCTGTGTGACACACGGCGCGGTGGGCTGCGCAGACGCGCGGAG 5943
QY 465 CTGGGCTGCG 524
DB 5944 GACTTACCGTCTCTGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6003

QY 525 GGCCTTGACGAGTGTTTGACACGCTGTTTGGCGCGCACTGACATCATGAGAGGCG 584
DB 6004 CCCTTCGGAGAGGTGAGGTCTATCCGCCGAAGGCCCGGAGCCGACGAGCCGACGCG 6063
QY 585 ACGAGAGCTGTGACCGCTTACACGAGACCTGAGAGCCACCCGATCTCCGACGCGCG 644
DB 6064 GGTGTGTGTGTCTGGCTCCGCAACGGAATGTCCGACCCGGCTTGAAGGCGCGAGAAC 6123
QY 645 CTGGCCATGTTCACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 704
DB 6124 GCGCGGAGAGTGTGCGAGGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6183
QY 705 GACTGTATCCCTTACGAGTGTGACGAGTGTGACGAGTGTGACGAGTGTGACGAGTGT 764
DB 6184 ATGCTGTACCGGGGTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6243
QY 765 AAGTC-----CTACTAGCGGAAAGAGGCGATGCGCGCAAGGACATGTGATGTG 818
DB 6244 AAGAGCTGTGAGTGTGCG 6303
QY 819 TCCATCATGCTGTGACCGCGCAAGCATGTGAGGCTGTGACCGGAGTGTGCTGTGTGAG 878
DB 6304 ACGACCGACACCGCGCGCAAGAGTGTGCGGTGTGAGTGTGCGCGCGCGCGCGCGCG 6363
QY 879 GCCGACCCCACTGCG 938
DB 6364 ATGCGCAAGGAGTGTGCG 6420
QY 939 TTCAGAGAGCGGCGCATCACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 998
DB 6421 GCCACGAGCG 6480
QY 999 GCGGT 1058
DB 6481 CGACCTTCAAGCGGT 6540
QY 1059 CTGGCGACG-----GCTATGAGCTGTTCACGGGACGCGCGCTG 1097
DB 6541 TTGCGCAAGCGCTGTGCG 6600
QY 1098 CCGCGCTGAGCTGTGACGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1157
DB 6601 GCGCGCTGTGCTGT 6660
QY 1158 ATGTGTGCTGCG 1217
DB 6661 GAGGTCTCAAGTACCG 6720
QY 1218 GCGGAGCG 1277
DB 6721 GTCACTTCTGCTGT 6780
QY 1278 ACCAGGAGGAGCGGAGGCGCGCGCATCACTGTGCGGTGTGCGCGCGCGCGCGCGCG 1337
DB 6781 GCCATGCGCATGCGCAAGT 6840
QY 1338 AACGCCAAGAGTGTATCAACCAAGATGACGGCCGGCGGAGCCCAAGTACGACTTTGTGAG 1397
DB 6841 CCCTTCGGAGAGTGTGAGTGTATCTGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCG 6900
QY 1398 ATCATGAGCTGCTGCCCCGGGGCTGTGGGCGGGGCGGCGCGCGCGCGCGCGCGCGCG 1457
DB 6901 CTGGGGGCGCGCTGTGCGGAGCATCTGCGGGGCGGAGAGTGTCTATCGGAGTGAACCTC 6960
QY 1458 GCCATCAGCA 1468
DB 6961 GCCATCGCGA 6971

RESULT 6
US-09-385-028-1
; Sequence 1, Application US/09385028

```

Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A. Aidoo
APPLICANT: Ashish S. Parashar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P57452US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces clavuligerus
US-09-385-028-1

Query Match 2.68; Score 63.4; DB 4; Length 15079;
Best Local Similarity 43.58; Pred. No. 0.00024;
Matches 475; Conservative 0; Mismatches 586; Indels 30; Gaps 3;

```

```

DB 8216 ATCGCTTCACCGGAGTATGCGCCGCGAGTACCCGATGAGAGATCCGCGAGCACTC 8275
QY 765 AAGTC-----CTACCTAGCGGAAAAAGAGGCGATCGCGCCCAAGACATGGTCAATG 818
DB 8276 AAGAGCTGTAGATGAGCCCGCCGCGGAGGCGGCTTCAGACCGCGCGCCGCGCATATG 8335
QY 819 TCCATCATGCTCTGACCGCGCAAGCATGCGAGCTGACCGGAGCTGCTTGTGTGAGC 878
DB 8336 AGGACCGACACCGCGCCCAAGAGGTCCGGGTCAAGGTGCGGAGGCGACCTCGTGGCG 8395
QY 879 GCGGACCCGACCTGCGCGCAGCTGAGACGATCATACCAACCGGTGAGGCAACATC 938
DB 8396 ATGCGCAAGGGGTGCGGATGCTGAGACCCGACATGCGAC---GCTGTGACCTTCTTC 8452
QY 939 TTCAGAGAGCGGCGGATCAACCTGCGCGAGCTCCCGAGGCGAGTGGACAAATCCATG 998
DB 8453 GCGACGAGCGCGCGGTGAGACCGCGCGAGACCGGCTCTCCGCGGCTCATGAGC 8512
QY 999 GCGGTGGGTCTGGGCGCGCGGTGCTGCTTCCGACACCGCGGCTCATGAGGCGGCG 1058
DB 8513 CGCACCTTCAACGCGGTGAGATGACACCGACACCTCCACAGCGACAGCGGCGTCTG 8572
QY 1059 CTGCGGACG-----GCTATGAGCTGTTCAAGGCGCGCGCTG 1097
DB 8573 TTGCGCAAGCGCTGCGCGCGCGAGGTGACGCTCGGGAGTTGAGAGGCGCTGCACAG 8632
QY 1098 CCGCGCTGAGCTGAGCGAGGTGCGCGGATGAGCGGCGATCAAGAGCAACATCACC 1157
DB 8633 GCGGCGCTGCGCTGCTGATCAAGGACATCGGACGAGCGGAGGCGCGCGCTGATC 8692
QY 1158 ATGTGTCGCGCGCGCGCGGTCCAAAGTTGAGAGCTGTGAAGCAGCGCGCGCGCGCG 1217
DB 8693 GAGGTCAAGTCAACGCGCGCGCGCGCGAGCGAGCGCCAGGCGGAGGTCCGCAAGCTC 8752
QY 1218 GCGGAGCGCGCGCGCGCGACACCGCGCGCGCTGCGCTGAGGACGCGCGCGCGCTTC 1277
DB 8753 GTCACTCCCGCTGTGTGTAAGACCGCGCTGACAGCTGCGACCCCACTGGGCGCGGTC 8812
QY 1278 ACCAGGAGAGCGGCGCGCGCGCATACACTGCGGTGCGCGCGCGCGCGCGCGCGCG 1337
DB 8813 GCCATGCGATCGGCAAGTGTCTGAGGACACCGAGATGACCGAGAGCGGCTGAGCATC 8872
QY 1338 AAGCCCAAGAGCTGATCAACAAAGATGACGCGCGGAGGCGCAAGTACGATTTGTGAG 1397
DB 8873 CGTTGCGGAGGTGAGGTCTATCCGCGGAGGCGCGGCGGCGACAGGCCGACGCGCG 8932
QY 1398 ATCATGCGCTGCGCGCGCGCGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCTTCA 1457
DB 8933 CTGCGGCGCGCGCGCTTCCGCGAGCATCTGCGGCGCGAGAGGTGATCTGCGGATG 8992
QY 1458 GCCATCACGCA 1468
DB 8993 GCCATCGCGGA 9003

RESULT 7
US-09-372-422A-11
; Sequence 11, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
; APPLICANT: Rudolph Jung
; APPLICANT: Francois Barrieu
; FILE REFERENCE: 0919
; CURRENT APPLICATION NUMBER: US/09/372,422A
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: US 60/098,692
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1217

```

```

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110) ... (974)
;
US-09-372-422A-11

```

```
Query Match      2.6%; Score 62.6; DB 4; Length 1217;
Best Local Similarity, 48.7%; Pred. No. 0.00019;
Matches 170; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
```

Qy	1004	GGGCTCGGGGCGCGGCTGCTGTTTGGGACCAACGCGGAGTGCATGAGAGCGGCGCTGG	106
Db	133	GGGGCCGAGGGGGGCGAGTTCTGGGCCAAGGACACACGACCCCTCCGCGCGCCCGCT	192
Qy	1064	CACGGCTATAGCTGTTCCAGGGGACGCGCTGCGCGGCTGAGCTGAGCGAGGTGG	1122
Db	193	GATCGACCCGGAGGAGCTCACCCAGTGTCACTGACCGCGGATCATCGCGAGTTGAT	252
Qy	1124	CGGCAATGACGSCATCAAGAGAGCCMAATCAATGATGCGCGCGCGGCTCCAACT	1183
Db	253	CGCCACGCTGCTTCTCTACATCACCGTGGCCACCGTATGGGTCAAGCACACAGC	312
Qy	1184	TGAGGAGCTGCTGAAGCAACCGGCGCGCGCGCGCCGAGGCGCGCGCACAGCAACC	124
Db	313	GGAGCGGTGGCGGTGGGCGCCCGAGCGGCGCGCGCGCGGCGGCGCATCTTGGGATGCG	372
Qy	1244	CGGGCGCTGGCTTGGGAGCGGCGCGCGGGGCTTCAACAGAGAGACGCGAGGGCGGCACT	1303
Db	373	TTGGGCGCTTGGCGGCGCATGATTTTCATCTCTGTACTAGCAACCGCGGATATTCAGGTTG	432
Qy	1304	CACACTGCGCGTGGCGCTTGGCCAAACGGGCTGGGGCAACGCAAGAACTTG	1352
Db	433	GCACATCAACCGCGCGTGAACCTTGGGCTGTTCTTGGCGAAGAGGTG	481

RESULT 8
US-09-003-287-7
; Sequence 7, Application US/09003287

```
Query Match      2.68; Score 61.8; DB 3; Length 735;
```

QY 875 GGAGCGGACCCCACTCTGCGCCAGCTGGACCACTCATCAACCGTGGAGCTGGGACA 934

Db 57 GGCCATCTGTGACATCTCTCTCTGGCAACCTGGCGCACCTGACCTCTACTCCGTGGAGAGCAT 116

OY 935 CATCTTCAAGGAGCGCGGCATCACTGGCGGAGTGGCCCGAGGCGAGTGGAGACATCC 994

Db 117 GCGCTTCCAGGCGCGCCGACACTGTGGCCGACGAGGCCAGGGCTTGTCTAAAGGCCCGCTT 176

OY 995 AATGGGCGTGGGCTGGGCGCGGCGGTGTTGCGGACACACGCGCGGTGTCATGGAGGC 105

Db	177	CTCCCGAGACCTACAACACATCATGCGGGTGTCTTACTATGGGGCACCGGATGCGAG	236
Oy	1055	GGCGCTGGCGAC-----GGCTATAGAGCTGTTACGGGAC--GCCGCTCCGGCCT	1105
Db	237	GAGCTCTCTCCCGAGCAGGCGCAAGAGACTCTCCCTCCACCTGGGCCCTCATCTTCC	296
Oy	1106	GAGCTAGCGAGGTGTGGCGGCATGACGGCATTAAGAGACCAATCATCATGTGTGCC	1165
Db	297	CCTTCACGAGGTGGGACCCGCGAGGCTTACTTCACTCCACCGCATGTCTTTCGACAT	356
Oy	1166	CGCGCCCGGGCTCAAGTTTGAAGAGCTGTGAAGACCGCGCGCGCGCGCGCGAGG	1225
Db	357	CTAGGGGGCGATCAAGCGGATGAGAGTCTCTAGAGTCTTGGCTCTTCAACGACAGGC	416
Oy	1226	CGCGCGCACCGCACCCCGGGCGCTGGAGCGAGCGCGCGGCGCTTACACAGCA	1285
Db	417	GGAAGCGGTGGCCGAGGCGCATATCCGCGACAGAGACGTGGCGGTGCACGGCAATCAC	476
Oy	1286	GGAGCGCAGGGGCGGCGATCACACTGCGCGGTGGCCGTGGCCCAAGGGCTGGGCACCGCCAA	1345
Db	477	CTTCTTGGGCGAGCTCATCCAGCTCTGCGCACCTTACGACAACTGGGCGCTTACGACGG	536
Oy	1346	GAACTGATCAACAAGATGCAAGCGCGCGGCGCAAGTACGACTTTGTGAGATCATGGC	1405
Db	537	CATGACGACATTCGGGCTCTTGGGTGGAGCAGACCAACCCGCACTTCAATCAACACCGCTT	596
Oy	1406	CTGCCCGCGGGCTGTGTGGCGCGCGCGCGCACCCCGCTCCACCGACAGAGCCATCAC	1465
Db	597	CCCGCGCACGGGCTGTGTGCTCTGTGTTCTGCTGACCGTGGCGAAGAGAGTCCACAA	656
Oy	1466	GCAGAGCGGCGAGGCGGCGCTGTACAACCTTGAACGAGAGTCCACGCTGCGCGCGACCA	1525
Db	657	GCCTGTGTGCACACCAACCATCCCGCAGTTTGACAAGACGATGAGAGGCCAACACCT	716
Oy	1526	CGAAGACCCGT	1536
Db	717	GATGAGAGCGGT	727

```

1  RESULT 9
2  US-07-945-283-1/c
3  ; Sequence 1, Application US/07945283
4  ; Patent No. 5352586
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Cheung, Andrew K.
7  ; APPLICANT: Wesley, Ronald D.
8  ; TITLE OF INVENTION: Pseudotables Virus Deletion Mutantes
9  ; TITLE OF INVENTION: Involving The EPO and LIT Genes
10 ; NUMBER OF SEQUENCES: 7
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: Curtis P. Ribando
13 ; STREET: 1815 No. 5352596th University Street
14 ; CITY: Peoria
15 ; STATE: IL
16 ; COUNTRY: USA
17 ; ZIP: 61604
18 ;
19 ; COMPUTER READABLE FORM:
20 ; MEDIUM TYPE: Floppy disk
21 ; COMPUTER: IBM PC compatible
22 ; OPERATING SYSTEM: PC-DOS/MS-DOS
23 ; SOFTWARE: Patentin Release #1.0, Version #1.25
24 ;
25 ; CURRENT APPLICATION DATA:
26 ; APPLICATION NUMBER: US/07/945,283
27 ; FILING DATE: 19920911
28 ; CLASSIFICATION: 424
29 ;
30 ; ATTORNEY/AGENT INFORMATION:
31 ; NAME: Ribando, Curtis P
32 ; REGISTRATION NUMBER: 27976
33 ; TELECOMMUNICATION INFORMATION:
34 ; TELEPHONE: 309-685-4011 ext. 513
35 ; TELEFAX: 309-685-4128
36 ;
37 ; INFORMATION FOR SEQ ID NO: 1:
38 ; SEQUENCE CHARACTERISTICS:

```


ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaeapiralis
US-08-510-646B-1

Query Match 2.64; Score 61.4; DB 3; Length 5392;
Best Local Similarity 43.34; Pred. No. 0.00048;
Matches 496; Conservative 0; Mismatches 636; Indels 13; Gaps 4;

384 GCCAAGCCCAAGAGCAGCAGCCGCGAGCAGCAGCTGCTGAGTGTCTCCGCGCTT 443
3900 GTGAGGGCCAGAGCAGCAGCTGACACAGAGGGCCCGCCAGCAGGGCTGATGTT 3959
444 CGTGTCTATTTCGAGAGCCTGGGCTGGCGCGGCGCCAGCAGCCCAAGCAGCTG 503
3960 GGCTACGCCAGCAGAGAGCCCTCGCTGATGGCTGCCCATGAGCTCGCCACCGC 4019
504 GCGAGGGCTCTCGCCCTCGCTTGAAGAGTGTTCAGACGCTTTGGCGCCGAC 563
4020 CTCTCGCGCGCTCACCGAGTCCGAGAGAGCGGACCGTCCCTACCTGCGCCGAC 4079
564 CTGACATCATGAGAGGAGGAGAGCTGCTGACCGCTCACCGAGCAGCTGAGAGCC 623
4080 ---GGCAGAGCCAGGTCACCATGAGTACAGAGGAGCCGCGCTGCGCTGAGACCC 4136
624 CACCCGACCTCCAGAGCCGCTGCCATGTTACAGAGCTGCTGCGCGCTGAGTCT 683
4137 GTGCTGCTCTCTCCAGAGCCGCGCCGACATCGACTCGGCTGCTCACCCTCGAC 4196
684 ATGTGAGAAATTTACCGGAGCTGATCCCTA-----CGTAGCAGCTGCAAGAGC 737
4197 ATCCGAGAGCAGCTGCTGAGAGAGCTCTGCGCCGACCTCGCCGAGAGAGGATCAGCTC 4256
738 CCCAGATGATGCTGCGGCGCCATGATCACTTACCTAGCGGAGAGAGAGGAGCTGCG 797
4257 GAGAGGAGCACTACCGCTGCTGCTGACACCCGCGCTTTCAGAGTGGCGGCGCG 4316
798 CCAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 857
4317 ATGGGCGAGCCGCGCTGAGCCGCGAGATCATGACAGTACGAGTGGCGGATGGCC 4376
858 CGGACCTGCTCTGAGAGCGCGAGCCGAGCCGCTGCGGAGGAGGAGCAGTATGATGAC 917
4377 CGCCACGCGGCTGCGGCTTCTCGGAGAGAGCCGCTCAGAGTGCAGCCGTTCCGCGCG 4436
918 ACCGTGAGAGTGGGCAATCTTCAAGAGAGCGGAGCATCAACTGCGGAGAGTCCCGAG 977
4437 TACCGGATGCGCTGGTCCCAAGAGTCTGCGCGCGGCTGCGCTGCGCTGCGAG 4496
978 GCGAGTGGAGCAATCCATGAGGCGTGGGCTGCGGCGCGGCTGCTGTTGCGCAGCAC 1037
4497 GTCCAGGTCGCTACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4556
1038 GCGGCTGATGAGAGGCGGCTGCGAGCGGCTATGAGCTGTTACAGGGGAGCGCGCTG 1097
4557 GGCACCGGAGCGTGGCGGAGAGCGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 4616
1098 CCGGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1157
4617 CGCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4676
1158 ATGATGCGCGCGCGGCTGCAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1216
4677 GCTTACGCGCACTTTCGCGCGGAGAGTGGCTTTCAGTGGAGGAGGAGGAGGAGGAG 4736
1217 CGCGGAGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1276
4737 CACCGGCTCAAGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4796
1277 CACGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336
4797 GCGCATGCTGTACACCGGCTGCTCATGCGCAGCAGCAGTATGATGATGATGATGATG 4853

1337 CACCGCAAGAGCTGATATCCAGAGATGAGGCGCGGAGGCGCAAGTACGATTTGTGA 1396
4854 GTTCGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4913
1397 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1456
4914 CGCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4973
1457 GGCATACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1516
4974 CGGCTCACCGCCGAGCTGATGAGGCGCGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 5033
1517 CGGCA 1521
5034 GCTCA 5038

RESULT 12
US-09-231-818-1
Sequence 1, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
NUMBER OF INVENTION: Coding For These Polypeptides And Their Use
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,818
FILING DATE:
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806,0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO

ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: S.pristinaespiralis
 US-09-231-818-1

Query Match 2.5%; Score 61.4; DB 4; Length 5392;

Best Local Similarity 43.3%; Pred. No. 0.00048;
 Matches 496; Conservative 0; Mismatches 636; Indels 13; Gaps 4;

384 GCCAAGCCCAAGAGCAGACCCGCGCAGACGCTGCGCGAGGCTCGCGCGCTT 443
 Db GTGAGGCGCAGAGCAGACGCTGACACAGAGGCGCGCGCAGCAGGCGCTGAT 3959
 QY 444 CGTGTCTATTTGCGAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 503
 Db GGTACGCGCAGCAGACGAGACCCCTCGCTATGCGCGCTGCGCGCGCGCGCG 4019
 QY 504 GCCGAGGCGCTCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCG 563
 Db CTCTCGCGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 4079
 QY 564 CTGACCTCATGAGAGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 623
 Db 4080 ---GGCAGAGCCAGGCTCAGCTACGAGGCGCGCGCGCGCGCGCGCGCG 4136
 QY 624 CACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
 Db 4137 GTGCTGTCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4196
 QY 684 ATGCTGAGAAATTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
 Db 4197 ATCCGCGAGCAGCTGCTGAGAGCGCTCGCGCGCGCGCGCGCGCGCG 4256
 QY 738 CCCAGATGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797
 Db 4257 GAGCGGACAACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4316
 QY 798 CCAAGAGCATGTGCTATGCTGCTATGCTGCTGCTGCTGCTGCTGCTG 857
 Db 4317 ATGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4376
 QY 858 CGGAGCTGCTTCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 917
 Db 4377 CGCGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4436
 QY 918 ACCGTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 977
 Db 4437 TAGCGATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4496
 QY 978 GCGAGTGGGCAATCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037
 Db 4497 GTCCAGGTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4556
 QY 1038 GCGGCTGTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1097
 Db 4557 GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4616
 QY 1098 CCGCGCTGAGCGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1157
 Db 4617 CG 4676
 QY 1158 ATGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1216
 Db 4677 GCGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4736
 QY 1217 CGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1276
 Db 4737 CACCGGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4796
 QY 1277 CACGAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1336
 Db 4797 GCGCATGCTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4853

QY 1337 CAAGCCCAAGAGCTGATACCAAGATCGAGCGCGCGCGCGCGCGCGCG 1396
 Db 4854 GTTCGGGATCACTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4913
 QY 1397 GATCAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1456
 Db 4914 CGCATCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4973
 QY 1457 GCGCATCAGCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1516
 Db 4974 CGGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5033
 QY 1517 CCGCA 1521
 Db 5034 GCTCA 5038

RESULT 13

US-08-804-439A-13
 Sequence 13, Application US/08804439A

Patent No. 6015565

GENERAL INFORMATION:

APPLICANT: Rose, Timothy M.

APPLICANT: Bosch, Wernix L.

TITLE OF INVENTION: GLYCOPROTEIN B OF THE HHV/KSHV

TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES

NUMBER OF SEQUENCES: 113

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Ste 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,439A

FILING DATE: February 21, 1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Halle, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09176/004001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

TELEX:

INFORMATION FOR SEQ. ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 2713 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-804-439A-13

Query Match 2.5%; Score 61; DB 3; Length 2713;

Best Local Similarity 43.2%; Pred. No. 0.00049;
 Matches 289; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 786 AAGGAGTGGCGCGCAAGCATGTGCTGCTATCATGCGCTGCAAGCGAG 845
 Db 2017 AACATACCATGTGAGGATACAGAGTTGCTCCCTGAGGCTACACCGCAG 2076
 QY 846 TCGAGGCTGAGCGCGCATGCTGTGTGAGCGCGCGCGCGCGCGCGCG 905
 Db 2077 ATCAAGGACAGCGCGCTGCTGAGACTACAGAGGTACAGCGCGCGCG 2136
 QY 906 CAGGTATACCAACCGTGAGCTGGCGCAATCTTCAAGAGCGCGCGCATCA 965


```

Db      2137 CTGCGCTTCCGCGACATCGACACGCTCATCCAGCCGACCCCAACCCCGCATGTTCCGG 2196
Qy      966 GACCTGCCCGAGGCGGAGTGGACATTCAAATGAGGGCTGGGGCTCGGGCCCGCGCTGGTG 1025
Db      2197 GGCGCTGGGGCGCTTCTTCGAGGGAGATGGGGACCTGGGGCGCGCGGCTCGGCAAGTGGTG 2256
Qy      1026 TTCGGACCAACCGCGGTGTATGAGAGCGCGCTGGCGACGGCTATGAGCTGTTACG 1085
Db      2257 ATGGGATCGTGGGGCGGCTGTATCGGCGGTGCGGGCTGTCTTCCTTCATGTCACAC 2316
Qy      1086 GCGACCGCGCTCGCGCGCGCTGAGCTTGAAGAGAGTGGCGCGGATGACCGCATCAAGAG 1145
Db      2317 CCTTTGGGGCGCTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2376
Qy      1146 ACCAACATCAACATGAGTGGCGCGCGCGCGCTGCAAGTTTGAAGAGTGTGTAAGACAC 1205
Db      2377 GCTTTGGCTATGCTATGCTGCGGTGCAAGACACCCCATGAAGCCCTGTACCGCTTAAC 2436
Qy      1206 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1265
Db      2437 ACCAAGAGCTCAAGAACCCCAACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2496
Qy      1266 GCGCGCGCGCTTCAACCAAGAGAGAGCGAGGGCGCGCATCACTGCGCGCGCGCGCG 1325
Db      2497 GACTTTGACGAGCGCAAGCTAGCCGAGCGCGCGGAGATGATGATGATGATGATGATG 2556
Qy      1326 AACGGGCTGGGCAACGCAAGAGCTGATCAACCAAGTGAAGCGCGCGCGCGCGCGCGCG 1385
Db      2557 TCTGCCATGAGAGCGGCAACGAGGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2616
Qy      1386 GACTTTGTGAGATGATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1445
Db      2617 GCCAAGGTCAACGACATGATGATGCGGAGCGCGCGCAACCACTACCGCAAGTTCCC 2676
Qy      1446 TCCACCGAC 1454
Db      2677 AACAAAGAC 2685

```

RESULT 14

```

US-08-720-229-13
Sequence 13, Application US/08720229
Patent No. 6023542
GENERAL INFORMATION:
APPLICANT: Rose, Timothy M.
APPLICANT: Bosch, Tamix L.
APPLICANT: Strand, Kurt
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV
NUMBER OF SEQUENCES: 100
CORRESPONDENCE ADDRESS:
ADDRESS: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720,229
FILING DATE: 26-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Schiff, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20002.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600

```

```

TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO.: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 2713 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-720-229-13

```

Query Match

```

2.5%; Score 61; DB 3; Length 2713;
Best Local Similarity 43.2%; Pred. No. 0.00049;
Matches 289; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

```

```

Qy      786 AAGGCGATCGCCCAAGACATGCTATGCTGTCTCATCATATGCTTGCACGGCGCAAGCG 845
Db      2017 AACATCACCATCTGAGAGATCAGAGATTTGCCCCCTGAGGTATACACCCGACAGG 2076
Qy      846 TCGAGGCTGACCGGAGCTGTTCTGTGTGAGCGCGGACCCGACCTGGCGGCGAGTGGAC 905
Db      2077 ATCAAGAGACGCGCTGCTGTGATACAGAGGTTCAGCGCTCGCAACCACTGACAGAC 2136
Qy      906 CAGTGTATCACCACGCTGAGCTGGGCAACATCTTCAAGAGCGCGGCAATCAACTGGCC 965
Db      2137 CTGCGCTTGGCGGATATGACACGGTATTCACGCGGAGCGCAAGCGCGCATGTTCCGG 2196
Qy      966 GAGTGTCCCGAGGCGAGTGGCAATCCATGGGCTGGGCTCGGGCGCGCGCGCTGCTG 1025
Db      2197 GCGCTGGCGCGCTTCTTCGAGGGAGATGGCGACTTGGGCGCGCGCGCTGGCAAGGTGG 2256
Qy      1026 TTCGGACCAACCGCGGTGTATGAGAGCGCGCTGGCGACGGCTATGAGCTGTTACG 1085
Db      2257 ATGGGATCGTGGGGCGGCTGTATGCGCGGTGTGCGCGGTGTCTCTTCATGTCACAC 2316
Qy      1086 GCGACCGCGCTGCGCGCGCTGAGCTGAGCGAGGTGCGCGGAGTGAACCGCATCAAGAG 1145
Db      2317 CCTTTGGGGCGCTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2376
Qy      1146 ACCAATCATCATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1205
Db      2377 GCTTTGCTATGATGATGCGGCTGACAGACCAACCCATGAAGGCTGTACCGCTTAAC 2436
Qy      1206 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1265
Db      2437 ACCAAGAGCTCAAGAACCCCAACCGGACCGCGAGCGCTCGGGAGGCGAGAGGCGCG 2496
Qy      1266 GCGCGCGGCTTCAACCAAGAGAGAGCGGCGGCGCATCACTGCGCGCTGGCGCC 1325
Db      2497 GACTTTGACGAGGCGCAAGCTAGCGCGGAGGCGCGGAGATGATGATGATGATGATG 2556
Qy      1326 AACGGGCTGGGCAACCCCAAGAGCTGATCAACCAAGATGACAGGCGCGGAGGCGCAAGTAC 1385
Db      2557 TCTGCCATGAGGCGGCAACCAAGGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2616
Qy      1386 GACTTTGTGAGATGATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1445
Db      2617 GCCAAGGTCAACGACATGATGATGCGGAGCGCGCGCAACCACTACCGCAAGTTCCC 2676
Qy      1446 TCCACCGAC 1454
Db      2677 AACAAAGAC 2685

```

RESULT 15

```

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6293328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

```

```

1 TITLE OF INVENTION: TUBERCULOSIS
2 FILE REFERENCE: 24366-20007.00
3 CURRENT APPLICATION NUMBER: US/09/103,840A
4 CURRENT FILING DATE: 1998-06-24
5 NUMBER OF SEQ ID NOS: 2
6 SOFTWARE: PatentIn Ver. 2.1
7 SEQ ID NO 2
8
9     LENGTH: 4403765
10
11     TYPE: DNA.
12
13     ORGANISM: Mycobacterium tuberculosis
14     FEATURE:
15
16     OTHER INFORMATION: CDC 1551
17
18     OTHER INFORMATION: "n" bases at various positions throughout the sequence
19
20     OTHER INFORMATION: represent a, t, c or g
21
22     US-09-103-840A-2

```

Query Match	2.5%	Score 60.4	DB 4	Length 4403765
Best Local Similarity	48.8%	Pred. No. 0.0041		
Matches 163	Conservative	0	Mismatches 171	Indels 0
			Gaps	0
QY	997	TGGGCGTGGGCTCGGGCCGCGCGTGCCTGTTTCGGCACCACCGGCGGCTGATCGAGGCGG	1056	
Db	2059204	TGGGCGGAGTTTGGGGGTTAGCGGTGGGATTCTCGGGGACCGCCCCCGCGCGCGCGGTG	2059145	
QY	1057	CGCTGCGCAGCGCCTATAGCTGTTTACAGGCGACGCGCTGCGCGCCTGAGCTTAGCG	1116	
Db	2059144	GTCGCGGCGGGGCGGCTGGGCTGTTTCGCGCTGCGTGGGCGCGGCGCGCGCGGCGGAGATCG	2059085	
QY	1117	AGGTGCGGGGATGAGACCGGCATCAAGAGACCAACATACCATAGTGTGCGCGCGCCGGGT	1176	
Db	2059084	GATTGTCGTGGAAACCGCGGTGTCGCGGGGGGTTCGGCGGGTTCGCGCTGCTTGAGGGACAG	2059025	
QY	1177	CCAAGTTTGAAGAGCTGCTGAAGACACGCGCGCGCGCGCGCGGCGGCGCGCCGACAG	1236	
Db	2059024	GCGGTGCGGGCGCGCGCGGGTGGGGTTCGAGGTTCACCTACCGGCGGTGCGCGCGGGCGGGCG	2058965	
QY	1237	GCACCCCGGGCCCGCTGCGCTGGGACGCGCGGCGCGGGCTTTCACCAACGAGACGCGACGG	1296	
Db	2058964	GCAACGCGCGGCTGCTGTGTAAGGCGCGCGCGGGCGGGCGCGGCGCGGACATCGCGGTTG	2058905	
QY	1297	GCGGCATCACACTGCGCGCTGGCGCGGTGGGCGCAACGG	1330	
Db	2058904	GCGCTACCGGAGTGGCGCGGCGCGCGCGGAGAACGG	2058871	

Search completed: June 8, 2003, 13:31:28
Job time : 2876 secs

This Page Blank (uspto)


```
QY 702 CCGGACCTGATCCCTTACCTGAGCAGCTGCAAGAGCCCCCAATGATCTGGCGGCATG 761
DB 736 GCCCATTGCAACATCAGCGCGGAGAGTGAACAACCTTGAAGAGATCTGACCAAG 795
QY 762 GTCAAGTCTTACCTAGCGGAAAAGAGGCGATCGCCCAAGGACATGCTCATGCTGTC 821
DB 796 CTGCAAGGCCCAAGTTCGCGCAACAAAGACCATCTGTTTCAAGCAGAGCAGCGCGCGACCC 855
QY 822 ATCATGCCCTGCAAGCGGCAAGCAGTGGAGCTGACCGCGCATGCTTCTGTGTGACGCC 881
DB 856 GAGATCGTGTATGCAAGCTTCAACTGGCGGAGTCTTCTTACTGCAACAGACCCAG 915
QY 882 GACCCACCTGCGGCAAGCTGAGCAAGTATCAGCAGGTGAGAGCTGGGCAATCTTC 941
DB 916 CTGTTCAACAGCATCTGGAACAACATCTGCGCCCAACCAACAGCAGCATTAC 975
QY 942 AAGAGCGCGCATCAACCTGCGGAGCTGCGGAGGCGAGTGGGCAATTCATGGG 1001
DB 976 CTGCGCTGCGCATCAAGCAGATCATCACTGGAGAGAGTGGGCAAGGCGCATGTAC 1035
QY 1002 GTGGGCTGGGCGCGCGC--GTGCTGTTGGCAACAAGCGGCTGTATGAGGCGGCG 1058
DB 1036 GCCCGCCCATCCGCGCGCAGATCCGCTGAGCAGCAACATCAAGCGCTGTGTGACC 1095
QY 1059 CTGCGCAGCGGCTTATGAGCTGTTCAAGGAGACCGCGCTGCGGCGCTGAGCCTGAGGAG 1118
DB 1096 CGGACGCGCGGCAAGAGATCAGCAACACAGAGTCTTCCGCGCGCGCGCGCGGAC 1155
QY 1119 GTGCGCGGATGAGCAGGATCAAGAGACCAACATCACTGATGTT--GCCCGCGCGG 1175
DB 1156 ATGCGCAACAATGCGGCGCAGAGCTGTATCAAGTACAGGTGTGAAGATCGAGCCCTG 1215
QY 1176 TCCAAAGTTTGAAGAGCTGTGAAGACACCGCGCGCGCGCGCGCGGCGCGCGGAC 1235
DB 1216 GCGGTGCGCCCAACCAAGCGCGCGCGGTGTGAGCGGCAAGCGCGCGGTGACC 1275
QY 1236 GCGACCGCGGCGCGCTGCGCTGGAGAGCGCGCGCGGCTTCAACAGCGAGAGCGGAG 1295
DB 1276 CTGGGCGCCATGTTCTTCTGGGCTTCTTGGGCGCGCGCGCGCAGCATGGGCGCGGAC 1335
QY 1296 GCGGCGATCACTGCGCGGTGCGGTGCGCAAGCGGCTGGCAACCGCAAGAGCTGATC 1355
DB 1336 CTGACCTGACCGTGA--GGCGCGCAGCTGTGAGCGGATCTGTCGAGCAGCAGAAC 1392
QY 1356 ACCAAGATGACGCGCGGCGAGCGCAAGTACGACTTTGTGAGATCATGCTTG 1408
DB 1393 AACCTGCTGCGCGCATGAGCGCCAGCAGCAGCAGCTGCTGAGCTGACCGTGTG 1445

RESULT 2
US-09-476-242-4
; Sequence 4, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: MARTIN, Eric
; APPLICANT: HARTOG, Karin
; FILE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-11e201
US-09-476-242-4
Query Match 2.9%; Score 69.8; DB 10; Length 2316;
Best Local Similarity 44.3%; Pred. No. 4e-09;
```

```
Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;
QY 462 ACCCTGGGCTTGGCGCGCGCCACACACCCCAAGAGCTGGCGAGGCGCTCCGCCG 521
DB 502 ACCCAAGGATCGCGCGCGCTGTGTGAGCACCCAGCTGCTCTTAAAGGCGCGCGGAG 561
QY 522 CTGGGCTTGAAGAGGTGTTTACAGCGGCTGTTTGGCGCGCACTGACATCATGAGAG 581
DB 562 GAGGCGTGTGTATCCGAGCGAGAACTTCAACCGACAGCCCAAGACCTCATCTGTGAG 621
QY 582 GCGAGAGCTGTGTACACCGCTTACCGAGCACTGAGAGCGCCACCGCATCTCGAGAG 641
DB 622 CTGAAGAGAGGTGAGATCAATGACACCCCGCCCAACAAACACCGCAAGAGCATC 681
QY 642 CCGCTGCCATGTTCAACAGCTGCTGCGCGCGCTGATGTGATGTGAGAAATTTAC 701
DB 682 ACCATCGCGCGCGCGCGCTTCTTACGCAACCGGAGATCATGCGCAATCCGCCAG 741
QY 702 CCGGACCTGATCCCTTACGTGAGCAGCTGCAAGAGCGCCAGATGATGCTGCGCGCATG 761
DB 742 GCCCATTGCAACATCAGCGCGAGAGAGTGAACAACACCTTGAAGCAGATCTGTACCAAG 801
QY 762 GTCAAGTCTTACCTAGCGGAAAAGAGGCGATCGCGCCAAAGACATGCTATGTGTC 821
DB 802 CTGCAAGGCCAGTTGGCGCAACAAGACCATCTGTTCAGCAGAGCAGCGCGCGACCC 861
QY 822 ATCATGCCCTGCAAGCGGCAAGCAGTGGAGCTGAGCGGAGCTGTTCTGTGTGACGCC 881
DB 862 GAGATGTGTATCACAGCTTCAATCGCGCGCGGAGTTTCTTACTGCAACAGCACCCAG 921
QY 882 GACCCACCTGCGCGAGCTGAGCAACATCATCAGCAGCTGAGAGCTGCGCAACATCTTC 941
DB 922 CTGTTTCAAGACATCTGGAACAACACATCGGCGCCCAACAACACCAAGCGCATCTAC 981
QY 942 AAGAGCGGCGGATCAACTGCGCGAGCTGCCGAGGCGAGTGGGCAATTCATGGGCG 1001
DB 982 CTGCGCTGCGCATCAAGAGATCATCAACCTGTGCAAGAGTGGGCAAGGCGCATGTAC 1041
QY 1002 GTGGGCTCGGCGCGCGC--GTGCTGTTGGGCAACCGCGGCGGTGTCTGAGAGCGGAG 1058
DB 1042 GCCCGCCCATCGCGCGCGAGATCGCTGCGCAGCAACATCAACCGCGCTGTGTGACC 1101
QY 1059 CTGCGCAGGCGCTATGAGCTGTTTCAAGGCGCAGCGCTGCGCGCTGAGCGTGAAGCGAG 1118
DB 1102 CGGATGCGGCGCAAGAGATCAAGCAACACACAGATCTTCCGCCCGCGCGCGCGAG 1161
QY 1119 GTGCGGCGATGAGCAGCATCAAGAGACCAACATCACTGATGTT--GCCCGCGCGG 1175
DB 1162 ATGCGGCAACAATGCGCGCAGCAGCTGTACAAAGTACAAAGTGTGAAGATCGAGCCCTG 1221
QY 1176 TCCAAAGTTTGAAGAGCTGTGAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGGAC 1235
DB 1222 GCGGTGCGCCCAACCAAGGCGCGCGCGGTGTGACAGCGGAGAGCGCGCGTGAACC 1281
QY 1236 GCGACCGCGGCGCGCTGCGCTGTGAGAGCGCGCGCGGCTTCAACAGCAGAGCGGAG 1295
DB 1282 CTGGGCGCATGTTCTTCTGGGCTTCTTGGGCGCGCGCGGAGCAGCATGGGCGCGGAC 1341
QY 1296 GCGGCGATCACTGCGCGGTGCGGTGCGCAAGCGGCTGGGCAACCGCAAGAGCTGATC 1355
DB 1342 CTGACCTGACCGTGA--GGCGCGCAGCTGTAGGCGGCTGTGAGCAGCAGAAC 1398
QY 1356 ACCAAGATGACGCGCGGCGAGCGCAAGTACGACTTTGTGAGATCATGCGCTG 1408
DB 1399 AACCTGCTGCGCGCATGAGCGCCAGCAGCAGCAGCAGCTGCTGAGCTGACCGTGTG 1451

RESULT 3
US-09-476-242-8
; Sequence 8, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
```

```

; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Thr202
US-09-476-242-8

```

```

Query Match      2.9%; Score 69.8; DB 10; Length 2316;
Best Local Similarity 44.3%; Pred. No. 4e-09;
Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

QY 462 ACCCTGGGCTGTGGCGCGGCGCCACACCCCAAGAGCTGGCCGAGGCTCCGCGC 521
DB 502 ACCCAGGAGATCGGCGCGCGCGCGGAGAGACCCAGCTGTGTAACGGCAGCTCGCGAG 561
QY 522 CTCGGCTTTGACGAGGTGTTGACAGCGCTGTTGGCGCGCAGCTGACCATCATGAGAG 581
DB 562 GAGGGCGTGTGTATCCGACAGCGAGAACTTACCGACAAAGCCCAAGACATCATGTGAC 621
QY 582 GCGAGGAGCTGTGTGACCGGCTTACCGACCATCTGAGAGCCCAACCGCATCTCGAGAG 641
DB 622 CTGAGAGAGAGCGTGTGAGATCACTGACCGCGCCCAACAAACCGCCCAAGAGATC 681
QY 642 CCGCTGCCCATGTTTCAACAGCTGTGCGCGCTGTGATCGTATGCTGAGAAATCTTAC 701
DB 682 ACCATGGGCGCGCGCGCGCGCTTCTAGCGACCGCGACATCATCGCGCATCTCGCGAG 741
QY 702 CCGGACCTGATCCCTTACGTGAGAGCTGACAGAGCCCGCATGATGTGGCGCGCATG 761
DB 742 GCCCATCTGCAACATCAAGCGCGAGAGAGTGAACAACCTCTGAAGCATGTGACCAAG 801
QY 762 GTCAATCTTACTAGGGAAGAGAGGATCGCGCGCAAGACATGTGATGTGTC 821
DB 802 CTGCAAGCGCGGATTTGCGCAACAAGACCATGTGTTCAAGCAGAGCAGCGCGCGAGCCC 861
QY 822 ATCATGCGCTGCAAGCGCAAGAGCTGAGAGCTGACCGCATGTGTTGTGAGAGCC 881
DB 862 GAGATGTGATGACAGCTTCAACTGCGCGGAGATTTCTTACTGCAACAGACCATCAG 921
QY 882 GACCCCACTGTGGCGCAGCTGACACAGTATCAACACCGTGAAGCTGGGCAATCTTTC 941
DB 922 CTGTTCAACAGCACTGTGAACAACATCATCGGCGCGCAACAACAGGACATCATCACC 981
QY 942 AAGAGCGCGGATCAACCTGTGGCGGAGCTGCGGAGGCGAGTGGACAATCCAATGAGC 1001
DB 982 CTGCGCTGCGGATCAAGCAGATCATCAACCGCTGAGAGAGTGGCAAGGCGCATGTAC 1041
QY 1002 GTGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
DB 1042 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1101
QY 1059 CTGCGCAAGGCTATGAGTGTTCACGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1118
DB 1102 CCGGACGCGCGGCAAGAGATCAGCAACACCGAGATCTTTCGCGCGCGCGCGCGCGCG 1161
QY 1119 GTGCGCGGATGAGCGGATCAAGAGACCAATCACTGATGTT--GCGCGCGCGCGCGCG 1175
DB 1162 ATGCGGCAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221
QY 1176 TCGAAGTTTGAAGAGCTGTGAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1235
DB 1222 GGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281
QY 1236 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295

```

```

DB 1282 CTGGGCGCATGTCTCTGGGCTTCTGTGGCGCGCGCGCGAGACCATGCGCGCGCGCAGC 1341
QY 1296 GGGCGCATCACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
DB 1342 CTGACCTTACCGTGA---GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1398
QY 1356 ACCAAGATCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
DB 1399 AACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1451

```

```

RESULT 4
US-09-476-242-5
; Sequence 5, Application US/09476242
; Patent No. US2002014683A1
; GENERAL INFORMATION:
; APPLICANT: BARRETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2322
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Ile201B
US-09-476-242-5

```

```

Query Match      2.9%; Score 69.8; DB 10; Length 2322;
Best Local Similarity 44.3%; Pred. No. 4e-09;
Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

QY 462 ACCCTGGGCTGTGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
DB 502 ACCCAGGAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 561
QY 522 CTCGGCTTTGACGAGGTGTTTGAACAGCTGTGTTGGCGCGCGCGCGCGCGCGCGCGCG 581
DB 562 GAGGGCGTGTGTATCCGACAGCGAGAACTTACCGCAACAGCGCAACATCATGTGAGAG 621
QY 582 GCGAGGAGCTGTGTGACCGCGCTTACCGAGCACCTGAGAGCCCAACCGCGCATCTCGAG 641
DB 622 CTGAAAGAGAGCGTGTGAGATCAACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 681
QY 642 CCGCTGCCCATGTTTCAACAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
DB 682 ACCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 741
QY 702 CCGGACCTGATCCCTTACGTGAGAGCTGCAAGAGCCCGCATGATGTGCGCGCGCATG 761
DB 742 GCCCATGTGAACATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 801
QY 762 GTCAATCTTACTAGGGAAGAGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 821
DB 802 CTGCAAGCGCGGATTTGCGCAACAAGACCATGTGTTCAACGACGAGCGCGCGCGCGCG 861
QY 822 ATCATGCGCTGCAAGCGCAAGAGCTGAGAGGCTGACCGGACCTGTTGTGTGAGCGCC 881
DB 862 GAGATGTGATGACAGCTTCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 921
QY 882 GACCCCACTGTGGCGCAGCTGACACAGTATCAACACCGTGAAGCTGGGCAACATCTTTC 941
DB 922 CTGTTCAACAGCACTGTGAACAACATCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 981
QY 942 AAGAGCGCGGATCAACCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1001
DB 982 CTGCGCTGCGCGCATCAAGCAGATCATCAACCGCTGTGAGAGAGTGGCAAGGCGCATGTAC 1041

```

```

QY 1002 GTGGGCTGGGCGCCGGC---GTGCTGTTGGGACCAACCGCGGTGTCAATGAGAGCGGCG 1058
DB 1042 GCGCCCGCCATCCGCGGCGGAGATCCGCTGAGCAGCAATCAACGGCCGTGCTGTGAC 1101
QY 1059 CTGCGCAGCGGCTATGAGCTGTTTACCGGACCGCCCTGCGGCTGAGCTGAGCGAG 1118
DB 1102 CGGACGCGCGGCAAGGAGATCAGCAACACCGAGATCTTCGCGCCCGCGCGCGAG 1161
QY 1119 GTGCGCGGATGAGCGGATCAAGAGACCAATCATCCATGGT---GCCCGCGCGGAG 1175
DB 1162 ATGGGACACAATGCGGACGAGCTGTATACAGTACAGGTGTTGAAGATGAGAGCCCTG 1221
QY 1176 TCCAAATTGAGAGCTGTGAAGACACCGCGCGCGCGCGCGCGAGCGCGCGAC 1235
DB 1222 GGGGTGCGCCCGCACCAAGGCGCAAGCGCGGTGTGAGCGCGCAAGCGCGCTGTACC 1281
QY 1236 GGCACCGCGCGCGCGCTGCGCTGGGACGCGCGCGCGCGCTTACACAGCGAGACGCGAG 1295
DB 1282 CTGGCGCCCATGTTCTTGGGCTTCTGCGCGCGCGCGCGCGACCATGGGCGCGCGAGC 1341
QY 1296 GCGCGCATCACTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
DB 1342 CTGACCTGACCGTGCA---GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1398
QY 1356 ACCAAGATGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
DB 1399 AACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1451

```

RESULT 5

```

US-09-476-242-6
: Sequence 6, Application US/09476242
: Patent No. US20020146683A1
: GENERAL INFORMATION:
: APPLICANT: BARNETT, Susan
: APPLICANT: HARTOG, Karin
: APPLICANT: MARTIN, Eric
: TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
: FILE REFERENCE: 1605.002
: CURRENT APPLICATION NUMBER: US/09/476,242
: CURRENT FILING DATE: 1999-12-30
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 6
: LENGTH: 2328
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Lys121-Val1200
US-09-476-242-6

```

Query Match 2.9%; Score 69.8; DB 10; Length 2328;

Best Local Similarity 44.3%; Pred. No. 4e-09;

Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

```

QY 462 ACCCTGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
DB 508 ACCCAGCGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
QY 522 CTCGGCTTTGAGAGGTGTTTGAACGCTGTTTGGCGCGCGCGCGCGCGCGCGCGCG 581
DB 568 GAGGCGGTGTGATCCGAGCGAGCACTTCAACGCAACGCGCAAGACCATATGCTGAG 627
QY 582 GCGACGAGCTGCTGACCGGCTTCAAGAGCACTGAGAGCGCGCGCGCGCGCGCGCG 641
DB 628 CTGAAGAGAGGTGTGATCACTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
QY 642 CCGCTGCCCATTTTCAACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
DB 688 ACCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
QY 702 CCGGACCTGATCCCTACGTGAGCAAGTGCAGAGAGCGCGCGCGCGCGCGCGCGCG 761

```

```

DB 748 GCCATCGCAATAGCGCGGAGAGTGAACCAACCTTGAACAGATGTGACCAAG 807
QY 762 GTCAAGTCTTACTTACGCGAAGAGGAGCATTCGCGCGCGCGCGCGCGCGCGCGCG 821
DB 808 CTGACAGGCGCGAGTTGGGCAACAGACATTCGTTTCAAGCAGAGAGCGCGCGCGCG 867
QY 822 ATCATGCCCTGACCGCGCAAGCATGCGAGGTGACCGCGAGTGTCTGTGTGACCGC 881
DB 868 GAGATGTGATGACAGAGTTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 927
QY 882 GACCCACCGCTGCGCGAGTGGACAGCTCATCAACCGCGCGCGCGCGCGCGCGCGCG 941
DB 928 CTGTTTCAACAGACCTTGAAACAAACATTCGCGCGCGCGCGCGCGCGCGCGCGCG 987
QY 942 AAGAGCGCGGATCAACCTGCGCGAGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCG 1001
DB 988 CTGCCCTGCGCATCAAGAGATTCATCAACCGCTGCGAGAGTGGCGCAAGCGCGCAT 1047
QY 1002 GTGGGCTGGGCGCGCGCG---GTGCTGTTGGCACACCGCGCGGTGTATGAGCGCG 1058
DB 1048 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1107
QY 1059 CTGCGCAGCGCGCTATGAGCTGTTTCAAGGACGCGCGCTGCGCGCGCGCGCGCGCG 1118
DB 1108 CGGACGCGCGCGCAAGAGATCAAGCAACACCGAGATTTTCCGCGCGCGCGCGCG 1167
QY 1119 GTGCGGAGTGAAGCGCATCAAGAGCAACATTCACCATGT---GCCCGCGCGCG 1175
DB 1168 ATGCGGACAACTGGCGCGAGCGAGCTGTCAAGTCAAGGTGTGAAGTTCAGCGCG 1227
QY 1176 TCCAAATTGAGAGCTGTGAAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1235
DB 1228 GGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287
QY 1236 GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
DB 1288 CTGGGCGCGCATGTTCTTGGGCTTCTTGGGCGCGCGCGCGCGCGCGCGCGCGCG 1347
QY 1296 GCGGCGATCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
DB 1348 CTGACCTGACCGTGCA---GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1404
QY 1356 ACCAAGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
DB 1405 AACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1457

```

RESULT 6

```

US-09-476-242-7
: Sequence 7, Application US/09476242
: Patent No. US20020146683A1
: GENERAL INFORMATION:
: APPLICANT: BARNETT, Susan
: APPLICANT: HARTOG, Karin
: APPLICANT: MARTIN, Eric
: TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
: FILE REFERENCE: 1605.002
: CURRENT APPLICATION NUMBER: US/09/476,242
: CURRENT FILING DATE: 1999-12-30
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 7
: LENGTH: 2334
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu122-Ser199
US-09-476-242-7

```

Query Match 2.9%; Score 69.8; DB 10; Length 2334;

Best Local Similarity 44.3%; Pred. No. 4e-09;

Matches 422; Conservative 0; Mismatches 522; Indels 9; Gaps 3;

Db 1324 CTGGGCGCCATGTTCTCTGGGCTTCTGGGCGCGCGGAGACCATGGGCGCCGACG 1383
QY 1296 GGGGGATCATCAGTCGCGCTGGCGCTGGCCAGCGGCTGGGCAACGCCAAGAGTATC 1355
Db 1384 CTGACCTTGACCGTGCA---GGCCCGCCAGCTCTGTAGCGGCACTCTGTGACGACGAGMC 1440
QY 1356 ACCAAGATGAGCGCGCGGAGGCGCAAGTACACTTTGTGAGATCATGGCCTG 1408
Db 1441 AACCTGTGGCGCCATCGAGGCCAGCAGCACTCTGTGACGCTGAG 1493

RESULT 8

US-09-808-880-1/c
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Belach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-808-880-1

Query March 2.7%: Score 64.6; DB 9; Length 50937;
Best Local Similarity 45.6%; Pred. No. 2.5e-07;
Matches 376; Conservative 0; Mismatches 419; Indels 30; Gaps 3;
QY 857 CCGGAGCTGTTCTGTGTGAAGCGCCGACCCCTGGCGCAGCTGACACAGTATC 916
Db 3774 CCGCCCTCGGGCGAGGTGTTCCCGCGGTGGGGCGGTGCTCTTACAGCGACCAT 3715
QY 917 CACCGTGAAGCTGGGCAACATTTTCAAGAGCGCGGATCAACTGGCCGAGTGGCCGA 976
Db 3714 CACCGTCCGGTCAACCGGACGCTACCGGATCGCGCCCGCGAGGGCGCAAGCTTA 3655
QY 977 GGGCGAGTGGGACATCAATGGGGGTGGGCTCGGGCGCGCGGTGTTTGGGACACC 1036
Db 3654 GTTCGAATCGAGCGCCAGAGACGCTTGGCCGCTGCGCTTGAAGCGGACGTAG 3595
QY 1037 CGGCGGTGTATGAGCGCGGCTGCGGACGCGCTATAGGTGTTTACGGGACCGGCT 1096
Db 3594 CAGCGGCCCATGCGGCTGGCGGGCGGAGACCAAGCTGCGCATCACGGGCGGGCCCT 3535
QY 1097 GCGGCGCTGAGCCCTGAGCGAGTGGCGGCGCATGAGCGGATCAAGAGAGACCAATCAG 1156
Db 3534 GCGGGAATGCGCCATGACCTTGTGAGCTGACGTGGTCACTCCGCGAGGCGCCAGAGGC 3475
QY 1157 CATGTGCGCGCGCGCGGCTCCAAATTGAGAGGTGCTGAAGCACCGCGCGCGCGCG 1216
Db 3474 GATGACCGGG---CGGTGTGATGCGCGCGAGCGGAGACCGGCGGTGTCTTGTGCGTA 3418
QY 1217 CGCCGAGCGCGCGCGGACGACCCCGCGCGCTGGCTGGGAGCGGCGCGCGGCTT 1276
Db 3417 CGAGCAGCGCTCGAGAGAGCGGAGACCGGACGTCGTCTCCGCGCGCGCGCGAGAGCA 3358
QY 1277 CACGACGAGAGCGGAGGGGGGCGATCATCTGGCGGTGGCGCGCGCGCA----- 1327

Db 3357 CTGATCAGCGCGTGGCGCGGCTAACCCAGACCTGTGTGTCTCAACAGGGCTC 3298
QY 1328 -----CGGCTGGGCGCAAGCGGAGGTGATCACAAGATGACGCGCG 1372
Db 3297 CTCCTTCAACATGCGCTGTGAGGAGACCGCGCGGTGTCTACCATGTGTACCGGG 3238
QY 1373 CGAGGCCAAGTATGACCTTTG---TGAATATATGGCTCTGCCCCGGGCTGTGTGGGGCG 1429
Db 3237 CGAGCGCGCGCGCGGAGGAGACACCGCCCTGTGTGGCGACGCGGACCCAGAGAGCGG 3178
QY 1430 GCGCGCGAGCCCGCTCAACCGCAAGGACCATCAGGAGCGGCGGCGGCTGT 1489
Db 3177 GTTACCCAGACCTTCTCCGCGCGAGAGGGGAGACGCTTGGCGGGGAGCGGAGGG 3118
QY 1490 CAACCTGAGCAGAAAGTCCACGCTGCGCGGACGACGAGAAACCGCTCATCCGAGCT 1549
Db 3117 CTACCGGAGATGATGACCAACTGAGACTACCGAGGCGATCTCTCGGCTACCGCTG 3058
QY 1550 GTACGACAGTACCTTGGAGAGCCGCTGGGCCACAAAGCGGACGAGCTGTGACACCA 1609
Db 3057 GTACGAAACGACGAGGCGTCCAGCCGCTGTCTTCCGCGCACGGGTTGTCTACACCTC 2998
QY 1610 CTACGTCGCGCGGCGGTGAGAGAGAGAGAGAGAAAGTGAAG 1654
Db 2997 CTTCGACTACCGGAGCTGAAGGTAGCGGAGCGGCGGAGCGGCGG 2953

RESULT 9

US-10-171-311-51
; Sequence 51, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kametkar, Shubhangi
; APPLICANT: Giat, Karen
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 15952
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-171-311-51

Query March 2.7%: Score 63.8; DB 9; Length 15952;
Best Local Similarity 49.8%; Pred. No. 3e-07;
Matches 161; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY 1142 GAGACCAACATCAACATGTCGCGCGCGCGGCTCCAAATTGAGAGCTGCTGAAGCA 1201
Db 8410 GAGGCGGCGCACCGGCTTGTATCAACCCGTCGCAACTGAGCTGTGCTGAGGA 8469
QY 1202 CCGGCGCGCGCGCGCGGAGGCGCGCGGACAGGACCGGCGCGGCTGTGGCTGGA 1261
Db 8470 GAGCTGTGCGCGCGGCTGTGTGGCGCGGAGATCAAGAGAAAGCTGTGTGGCGGAGCG 8529

Db 1231 GTGGCCCCACCAAGGCCAAGCGCGGTGTGACAGCGCGGAGAAAGCGCGCGGTGACCTG 1290

QY 1239 ACCCCGCGGCGCTGGCTGGAGCGCGCGGCGCTTACGACGAGAGCGGAGGCGC 1298
DB 1291 GGGGCGCATGTTCTGGGCTTCTGGGGCGCGCGGAGACCAATGAGCGCCGCGAGCTTG 1350
QY 1299 GGCATCACATGCGCGCGCTGGCGGCAAGCGCTGGGACCCCAAGAAAGCTGATCAC 1358
DB 1351 ACCCTGACCGGTGA---GGCGCGCGAGCTGCTAGCGCGCATGCTGACGACGACAAAC 1407
QY 1359 AAGATGCAAGCGCGCGAGCGCAAGTACGACTTTGTGAGATCATGCGCTG 1408
DB 1408 CTGCTGGCGCGCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1457

RESULT 12
US-09-476-242-10
; Sequence 10, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 2541
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Arg426-Gly431
US-09-476-242-10

Query Match 2.7%; Score 63.6; DB 10; Length 2541;
Best Local Similarity 43.9%; Pred. No. 2e-07;
Matches 417; Conservative 0; Mismatches 524; Indels 9; Gaps 3;

QY 462 ACCCTGGGCGCTGGCG 521
DB 733 ACCCAAGCGCATCG 792
QY 522 CTCGGCTTTGACGAGGCTTTGACACGCTGTTGGCGCGCGCGCGCGCGCGCGCGCGCG 581
DB 793 GAGGCGGTGGTATCCGACGAGGAGAACTTACCGACAAACCGCAAGACATCATCTGTCAG 852
QY 582 GCGAGCGAGCTGCTGCAACCGCTCAACGAGCACTGAGGCGCGCGCGCGCGCGCGCGCG 641
DB 853 CTGAAGGAGAGCGGTGAGATCACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912
QY 642 CCGCTGCCCATGTTTCAACCAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
DB 913 ACGATCGGCG 972
QY 702 CCGGAGCTGATCCCTTACGTAGAGCAAGCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCG 761
DB 973 GCGCACTGCAACATCAAGGCGGAGAGTGAACAAACCGTGAAGAGAGATGTGACCAAG 1032
QY 762 GTCAAGTCTCTCTAGCGGAAAGAGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 821
DB 1033 CTGCAAGCG 1089
QY 822 ATCATGCGCTGACCGGAGAGAGTGGAGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 881
DB 1090 CCGGAGATGTGATGACAGGTTCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1149
QY 882 GACCCCAAGCTGCG 941
DB 1150 CAGCTGTTCAACAGCACTGGAACAAACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1209
QY 942 AAGAGCGCGCGCGCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1001

DB 1210 ACCCTGCCCTGCGCGCATCAAGACATCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1269
QY 1002 GTGGGCTGGGCG 1061
DB 1270 CCCCCATCG 1329
QY 1062 CCGACGCGCTATGAGCTGTTTCAAGCGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1121
DB 1330 GACG 1389
QY 1122 CCGGCGCATGAGCGCGCATCAAGAGACCAACATCAACATGTT---GCGCGCGCGCGCGCG 1178
DB 1390 CCGGACCACTGCG 1449
QY 1179 AAGTTGAGAGAGCTCTGAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1238
DB 1450 GTGGCG 1509
QY 1239 ACCCG 1298
DB 1510 GCGGCGCATGTTCTTGGGCTTCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1569
QY 1299 GGCATCACACTGCG 1358
DB 1570 ACCCTGACCGTGCA---GGCG 1626
QY 1359 AAGATGCAAGCG 1408
DB 1627 CTGCTGGCGCGCATGAGCG 1676

RESULT 13
US-09-476-242-22
; Sequence 22, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2298
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-11e201;
US-09-476-242-22

Query Match 2.6%; Score 62.6; DB 10; Length 2298;
Best Local Similarity 44.5%; Pred. No. 3.7e-07;
Matches 426; Conservative 0; Mismatches 519; Indels 12; Gaps 4;

QY 458 CGAGACCTGGGCGCTGCG 517
DB 483 CTGAGACGACCGTGCAGTGCACCGCAATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542
QY 518 CCGCGCTCGGTTTGAAGAGGTGTTTGAACAGCGCTGTTGGCGCGCGCGCGCGCGCGCGCG 577
DB 543 GAAAGGAGCGCTGCG 602
QY 578 GAGAGCGAGAGCTGCTGACCG 637
DB 603 CAAGACCATATGCTGCGAGGTGAAGAGAGCGGTGAGATCA---CTGACCGCGCGCGCGCG 659
QY 638 CGAGCGCGCGCGCATGTTTCAACAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 697
DB 660 CAACACACCGCGAGAGCATCAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 719

```

QY 698 TTACCGGACCTGATCCCTACGTGACAGCTGCAAGAGCCGCCAGATGATGCTGGCGG 757
Db 720 CATCATCGGGGACATCGCGAGGCCCACTGCAACATCAGCGGAGAAAGTGAACAACAC 779
QY 758 CATGTCAAGTCTTACTTACCGGAAAAAGAAAGGAGATGCGCCCAAGACATGTCTATGT 817
Db 780 CCGAAGCAGATGTGACCAAGCTGCAAGGCCAGTTGTGGCAAGAACATCGTTGTTCA 839
QY 818 GTTCATCATGCTCTGACCGCGCAAGCATGTGAGAGCTGACCGGACCTGTTCTGTGGA 877
Db 840 GCGAGACGAGGGGCGGACCCCGAGATCGTATGACACAGTTCACTGCGCGCGGAGTT 899
QY 878 CGCGGACCCGACCTGCGCGAGCTGGAACAAGTATCACCACGCTGAGAGCTGGGCAACAT 937
Db 900 CTTCCTACTGCAAGACGACCGAGCTGTTCAACAGACACTGTGAAACAACATCGGCCCA 959
QY 938 CTTCAAGAGAGCGGGGATCAACTGTGCGGAGCTGCCGAGAGGAGATGGGACATTCAT 997
Db 960 CAACACCAACGGGACCATCACTCTGCTGCGCATCAAGCATCACTGCGCGCGCAT 1019
QY 998 GGGCGTGGGCTCGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1054
Db 1020 GTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1079
QY 1055 GGGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1114
Db 1080 GACCGCGGAGGGGCGGAGGATCAAGCAACACCGAGATCTTCGCGCGCGCGCGCGG 1139
QY 1115 CGAGGTGCGGGGATGAGCGGCGCATCAAGAGACCAACATACCATGT--GCCCGCGG 1171
Db 1140 CGACATGCGGCAACATGCGCGGCGAGCGGCTGTCAAGTACAGTGTGTGAAGATTCAGCC 1199
QY 1172 CGGGTCGAAGTTTGAGAGGCTGCTGAAGACCGCGCGCGCGCGCGCGCGCGCGCGG 1231
Db 1200 CTTGGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1259
QY 1232 GCAAGGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1291
Db 1260 GACCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1319
QY 1292 CAGGGGCGGATCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1351
Db 1320 CAGCCTGACCTTACCGTGCA---GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1376
QY 1352 GATCACCAAGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1408
Db 1377 GAACAACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1433

```

RESULT 14

US-09-476-242-23

Sequence 23, Application US/09476242

Patent No. US2002014683A1

GENERAL INFORMATION:

APPLICANT: BARNETT, Susan

APPLICANT: HARTOG, Karin

APPLICANT: MARTIN, Eric

TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES

FILE REFERENCE: 1605.002

CURRENT APPLICATION NUMBER: US/09/476.242

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 23

LENGTH: 2298

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:

OTHER INFORMATION: Val120-Ile201B, Ile424-Ala433

US-09-476-242-23

Query Match 2.6%; Score 62.6; DB 10; Length 2298;
 Best Local Similarity 44.5%; Pred. No. 3.7e-07;
 Matches 426; Conservative 0; Mismatches 519; Indels 12; Gaps 4;

```

QY 458 CGAGACCCCTGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 517
Db 483 CGTGAGCAGCTGTGACATGCAACCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGG 542
QY 518 CCGCCTCGGCTTTGACGAGGTGTTGACACGCTGTTTGCGCGCGCGCGCGCGCGG 577
Db 543 GAAAGCGACCTGTGCGGAGAGGGGCGTGTGATCCGACGAGAACTTACCGACAAACG 602
QY 578 GGAAGCGACGAGCTGTGACCGCGCTCAACCGACACTGAGAGCGCGCGCGCGCGG 637
Db 603 CAAGACCATCATGTGTGACGTAAGAGAGCGGTGAGATCAA---CTGACCGCGCGCA 659
QY 638 CGAGCGGCTGCCCATGTTTACCAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGG 697
Db 660 CAACAAACCCCGCAAGAGCATCACTGGCGCGCGCGCGCGCGCGCGCGCGCGG 719
QY 698 TTACCGGACCTGATCCCTACGTGACAGCTGCAAGAGCCGCCAGATGATGCTGGCGG 757
Db 720 CATCATCGGGGACATCGCGAGGCCCACTGCAACATCAGCGGAGAAAGTGAACAACAC 779
QY 758 CATGTCAAGTCTTACTTACCGGAAAAAGAAAGGAGATGCGCCCAAGACATGTCTATGT 817
Db 780 CCGAAGCAGATGTGACCAAGCTGCAAGGCCAGTTGTGGCAAGAACATCGTTGTTCA 839
QY 818 GTTCATCATGCTCTGACCGCGCAAGCATGTGAGAGCTGACCGGACCTGTTCTGTGGA 877
Db 840 GCGAGACGAGGGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 899
QY 878 CGCGGACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 937
Db 900 CTTCTACTGCAAGACGACCGAGCTTCAACAGCACTTGAACACCAACATCGCGCGG 959
QY 938 CTTCAAGAGAGCGGGGATCACTTGGCGCGAGCTGCCGAGAGGAGATGGGACATTCAT 997
Db 960 CAACACCAACGGGACCATCACTCTGCTGCGCATCAAGCATCACTGCGCGCGCAT 1019
QY 998 GGGCGTGGGCTCGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1054
Db 1020 GTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1079
QY 1055 GGGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1114
Db 1080 GACCGCGGAGGGGCGGAGGATCAAGCAACACCGAGATCTTCGCGCGCGCGCGG 1139
QY 1115 CGAGGTGCGGGGATGAGCGGCGCATCAAGAGACCAACATACCATGT--GCCCGCGG 1171
Db 1140 CGACATGCGGCAACATGCGCGGCGAGCGGCTGTCAAGTACAGTGTGTGAAGATTCAGCC 1199
QY 1172 CGGGTCGAAGTTTGAGAGGCTGCTGAAGACCGCGCGCGCGCGCGCGCGCGCGG 1231
Db 1200 CTTGGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1259
QY 1232 GCAAGGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1291
Db 1260 GACCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1319
QY 1292 CAGGGGCGGATCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1351
Db 1320 CAGCCTGACCTTACCGTGCA---GGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1376
QY 1352 GATCACCAAGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1408
Db 1377 GAACAACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1433

```

RESULT 15

US-09-476-242-24

Sequence 24, Application US/09476242

Patent No. US2002014683A1

```

1 GENERAL INFORMATION:
2 APPLICANT: BARNETT, Susan
3 APPLICANT: HARTOG, Karin
4 APPLICANT: MARTIN, Eric
5 TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
6 FILE REFERENCE: 1605.002
7 CURRENT APPLICATION NUMBER: US/09/476,242
8 CURRENT FILING DATE: 1999-12-30
9 NUMBER OF SEQ ID NOS: 26
10 SOFTWARE: PatentIn Ver. 2.0
11 SEQ ID NO 24
12 LENGTH: 2298
13 TYPE: DNA
14 ORGANISM: Artificial Sequence
15 FEATURE:
16 OTHER INFORMATION: Description of Artificial Sequence: Val120-Thr202,
17 OTHER INFORMATION: 116424-A16433
18 US-09-476-242-24

```

Query Match	2.6%	Score 62.6;	DB 10;	Length 2298;
Best Local Similarity	44.5%;	Pred. No. 3.7e-07;		
Matches 426;	Conservative	0;	Mismatches 519;	Indels 12;
				Gaps 4;

458 CGAGACCTTGGGGCCTGGGGCGGGGGGACCAACACCCCAAGAGAGTGGCCGAGGGCCTCCG 517
483 CGTGAGCAACCTGCTAGCTGACCCACGGGATTCGGCCCTGGTGGAGCACCCAGCTGGCT 542
518 CCGCTCGGGCTTTGACGAGGTGTTGACACGCTCTTTGGCGGCCCACTGACATCATGA 577
543 GAACGGCGAGCTGGCCGAGGAGGGCGTGGTATCCGACGGCAGAACTTCACTCGACCAACGC 602
578 GGAGGGCAcCCAGCTGCTGCACCCGCTTACCAGAGCACTGGAGGCCACCCGCACTCCGA 637
603 CAAGCCATCATCGCTGACGCTGAGAGAGAGCGTGAATCAA---CTGCATCCGCCCCCAA 659
638 CGAGCGCGTGGCCATGTTCACCAGCTGCTGGCCCCCGGCTGATTCGCTATGCTGAGAAATC 697
660 CAACAACAACCCGACAGAGCATTCACATCGGCCCGGGCGGGCTTTCTACGCCACCGGCGA 719
698 TTACCCGGACCTGATTCCTCTTACGTGAGACGCTGCACAGAGCCCCCAGATGATGCTGGGGC 757
720 CATCATCGCGCGCATTCGCGCAGGGCCCACTGCACATTCAGCGCGAGAAAGTGGAAACAAC 779
758 CATGATCAAGTCTTACTTACGGGAAAGAAAGGGGCATCGCGCAAAAGACATGATCATGT 817
780 CTTGAAGCAGATCTGAGCCAAAGCTGCAGGCCAAGTTCCGCAACAAGACATCTGTCTCAA 839
818 GTTCATCATGCTCCCTGCACGCGCAGACAGTCCGAGGCGTACCGCGCATGCGATTCTGTGGA 877
840 GCAGAGCAGCGCGCGCGGACCCCGAGATGGTATGACAGCTTCAACTCGCGCGGCGAGTT 899
878 GCGCCGACCCCACTCTGGCGCTGAGTGCATCATCACACCTGTGAGAGCTGGGCACAT 937
900 CTTTACTCTGCACACACCCAGCTTTTCAACAGACCTGTGAAACAACACCATTCGGCCCCA 959
938 CTTCAAGGAGCGCGGATCAACTGTAGCGAGCTGCCGAGGGCGAGTGGGCAATTCAT 997
960 CAACACCAAGCGCACCATTAACCTTCCCTTCGCGCATCAAGAGATCATTCGGGGGGCCAT 1019
998 GGGCGTGGGCTCGGGGCGCGGC- -GTGCTGTTCCGACACACCGGCGGTGTTCATGAGGC 1054
1020 GTACGCCCCCCCATCCGGCGCAGATCCGCTGTGCAGACCAATCACCGGCTGTGCT 1079
1055 GCGCGTGGCCACGGCTTATGACTGTGTTCACGGGCAAGCGCGCGCGGCTGAGCTGAG 1114
1080 GACCCCGCAGCGCGGGAGAAAGATCAGCAACACACCGAGATCTTTCGGCCCGCGCGG 1139
1115 CGAGGTGCACCGCATGAGCGGATCAAGAGAGCAACAATCACCATGAT---GCCCGCC 1171
1140 CGACATCGCGCGAACACTGGCGGACGAGCTGTACAAAGTACAAAGTGTGGAAGATGAGCC 1199
1172 CGGGTCCAAATTGAGAGCTGTGAAGACCGCGCGCGCGCGCGCGCGGAGCCGCGC 1231

Db	1200	CCTGGGCGTGCCCCCAACAAAGCCAAAGCCGCGGTGAGGCGGAGAAAGCGCGCGGT	1259
Qy	1232	GCACGCGACCCCCCGGCGCGCTGTGCTGGAAAGCGCGGCGCGGCGCTTTACACACGACGAGACGG	1291
Db	1260	GACCTGTGGGCGCCATGTTCTTCGGGCTTCTGTGGGCGCGCGCCGACACACATGTGGCGCGCCG	1319
Qy	1292	CAGGGGGCGGATCACTGCGCGCGTGGGCGGTGGCCAAAGCGGCTGGGCAACGCGCAAGAAAGCT	1351
Db	1320	CAGCTGACCTTGACCGTGCA--GGCCCGCCAGCTGCTGTAGCGGCAATGTGACAGACGA	1376
Qy	1352	GATCACCAAGATGACAGCGCGGCGAGGCCAAGTACGACTTTGTGAGAGCATGGCGCTG	1408
Db	1377	GAAACACTGTGTGGCGCGCATGACAGGCCACACAGACACTGTGTGACGCTGACCGGTGTG	1433

Search completed: June 8, 2003, 14:36:16
Job time : 383 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2003, 09:51:01 ; Search time 3028 Seconds
(without alignments)
12831.233 Million cell updates/sec

Title: AJ012098

Perfect score: 2399
Sequence: 1 ATCTTACATGACACACAA.....TGTAGCAGCACTTCTGAG 2399

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gse:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pln:*
21: em_gse_vrt:*
22: em_gse_fun:*
23: em_gse_mam:*
24: em_gse_mus:*
25: em_gse_oher:*
26: em_gse_pro:*
27: em_gse_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649	27.1	657	13	BI717904 1031022F1
2	637.4	26.6	648	13	BI532215 1024119G1
3	637	26.6	644	13	BM003317 1031109G0
4	614	25.6	622	13	BI716102 1031001E0
5	612.8	25.5	617	13	BI999849 1031079E0
6	596.4	24.9	601	12	BE724427 894076B10

7	556.4	23.2	558	10	AV389547
8	545	22.7	602	12	BG48385
9	543.8	22.7	599	13	BI716925
10	539.8	22.5	621	12	BG48384
11	536.4	22.4	538	10	AV396217
12	499	20.8	499	10	AV396237
13	498	20.8	498	10	AV386674
14	477.6	19.9	550	13	BI717903
15	476	19.8	631	13	BI532214
16	474.4	19.8	550	10	BE352293
17	471.2	19.6	541	10	BE351941
18	468	19.5	549	12	BE724426
19	464.8	19.4	490	10	BE352300
20	456.8	19.0	481	12	BE726557
21	455.4	19.0	717	12	BG48383
22	453.4	18.9	463	10	AV390537
23	446.2	18.6	456	13	BM003316
24	442.4	18.4	456	10	AV394544
25	422.4	17.6	509	13	BI994221
26	421	17.5	458	10	BE237741
27	414	17.3	415	10	AV639539
28	397.8	16.6	762	12	BF861236
29	384	16.0	385	12	BE725875
30	377	15.7	377	10	AV641632
31	374.8	15.6	393	10	AV389590
32	366.2	15.3	658	13	BI726441
33	362.8	15.1	436	10	AV634899
34	354	14.8	354	10	AV631034
35	352.8	14.7	641	13	BI875012
36	347.4	14.5	362	10	AW721316
37	333.2	13.9	585	10	AV389350
38	332	13.8	353	13	BI994585
39	312.8	13.0	482	10	AV395481
40	301	12.5	549	10	BE337478
41	292.4	12.2	308	13	BI725565
42	275.4	11.5	326	12	BE761208
43	268.4	11.2	576	13	BI529458
44	251.6	10.5	504	10	BE352294
45	235.2	9.8	505	10	BE452899

ALIGNMENTS

RESULT 1
LOCUS BI717904 657 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031022F12.Y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BI717904
VERSION BI717904.1 GI:15693599

KEYWORDS

EST.

SOURCE

Chlamydomonas reinhardtii.

ORGANISM

Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE

1 (bases 1 to 657)
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,

AUTHORS

P., McDerroet, J.P., Shrager, J., Sillitow, C. and Stern, D.

TITLE

Analyses of the Chlamydomonas reinhardtii Genome: A Model.

JOURNAL

Unpublished (2001)

COMMENT

Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chause@duke.edu

FEATURES

Location/Qualifiers
1..657
/organism="Chlamydomonas reinhardtii"

Db	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
Db	BM003317	644 bp mRNA linear EST 25-OCT-2001	BM003317										
Qy		1031109G06.y1 C. reinhardtii CC-1690, Stress II (normalized),											
Db		Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.	BM003317										
Qy		BM003317.1 GI:16438097											
Db		EST.											
Qy		Chlamydomonas reinhardtii.											
Db		Chlamydomonas reinhardtii.											
Qy		Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;											
Db		Chlamydomonas.											
Qy		1 (bases 1 to 644)											
Db		Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,											
Qy		P., McDermott, J.P., Shager, D., Silflow, C. and Stern, D.											
Db		Analyses of the Chlamydomonas reinhardtii Genome: A Model,											
Qy		Unicellular System for Analyzing Gene Function and Regulation in											
Db		Vascular Plants. Project: 1031											
Qy		Unpublished (2001)											
Db		Contact: Charles Hauser											
Qy		DCMB Box 91000											
Db		Duke University											
Qy		Durham, NC 27708-1000											
Db		Tel: 919 613 8159											
Qy		Fax: 919 613 8177											
Db		Email: chauser@duke.edu.											
Qy		Location/Qualifiers											
Db		1..644											
Qy		/organism="Chlamydomonas reinhardtii"											
Db		/strain="CC-1690 wild type mt+ 21gr"											
Qy		/db_xref="taxon:3055"											
Db		/clone_11b="C. reinhardtii CC-1690, Stress II (normalized											
Qy), Lambda Zap II"											
Db		/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:											
Qy		XhoI; Stress condition II library, constructed by John											
Db		Davies and Jeffrey McDermott, combines cDNAs from CC-1690											
Qy		cells grown to mid-log phase in TAP (NH4+ - containing)											
Db		and shifted to TAP - NO3- (24hrs); H2 production											
Qy		conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant											
Db		Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +											
Qy		sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).											
Db		PolyA mRNA was purified from each sample, pooled and cDNA											
Qy		synthesized. The cDNA was directionally cloned into Lambda											
Db		Zap II (Stratagene) in the EcoRI (5') and XhoI (3')											
Qy		sites. pBluescript II SK- plasmids were excised from the											

ORIGIN	BASE COUNT	118 a	213 c	230 g	83 t
Query Match	26.6%;	Score 637;	DB 13;	Length 644;	
Best Local Similarity	100.0%;	Pred. No. 5,4e-119;			
Matches 637;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	755	GGCCATGATCAAGTCTTACCTTAGCGGAAAGAGAGGCATCGCGCCAAAGACATGTGTCAT	814		
Db	8	GGCCATGATCAAGTCTTACCTTAGCGGAAAGAGAGGCATCGCGCCAAAGACATGTGTCAT	67		
OY	815	GGTCTTCATCATYGCCTCTGCACGCGCAGACAGTCSGAGGCTGACCGCGCACTGGTCTGTGT	874		
Db	68	GGTCTTCATCATGCTCTGCACGCGCAGACAGTCSGAGGCTGACCGCGCACTGGTCTGTGT	127		
OY	875	GGACGCGGACCCCAACCCCTGCGCGAGCTGGACACAGTCATCAACCAACCGTGAAGCTGGGCAA	934		
Db	128	GGACGCGGACCCCAACCCCTGCGCGAGCTGGACACAGTCATCAACCAACCGTGAAGCTGGGCAA	187		
OY	935	CATCTTCAGAGAGCGCGGCATCAACCTGTGCGGAGCTGCGGAGGGCCGAGTGGGCAATCC	994		
Db	188	CATCTTCAGAGAGCGCGGCATCAACCTGTGCGGAGCTGCGGAGGGCCGAGTGGGCAATCC	247		
OY	995	AATGGGCGTGGGCTCGGGCGCGCGCGCTGTGGGACCAACCGGCGGTGCATGAGAGGC	1054		
Db	248	AATGGGCGTGGGCTCGGGCGCGCGCGCTGTGGGACCAACCGGCGGTGCATGAGAGGC	307		
OY	1055	GGCGCTGGCGACGCGCCTATGAGCTGTTACAGCGGACGCGCTGCGCGCTGAGCCTGAG	1114		
Db	308	GGCGCTGGCGACGCGCCTATGAGCTGTTACAGCGGACGCGCTGCGCGCTGAGCCTGAG	367		
OY	1115	CGAGGTGCGCGGCATGAGCGGCATCAAGAGACCAACATCAATCATGTGTCCCGCGCCGG	1174		
Db	368	CGAGGTGCGCGGCATGAGCGGCATCAAGAGACCAACATCAATCATGTGTCCCGCGCCGG	427		
OY	1175	GTCCAAGTTGAGAGCTGTGAAGCAACCGGCGCGCGCGCGCCGACAGGCGCGCGCA	1234		
Db	428	GTCCAAGTTGAGAGCTGTGAAGCAACCGGCGCGCGCGCGCGCCGACAGGCGCGCGCA	487		
OY	1235	CGGACACCCCGGCGCGCTGTGGAGACGCGCGCGCGGCTTTCACAGCAGAGACGCGAG	1294		
Db	488	CGGACACCCCGGCGCGCTGTGGAGACGCGCGCGCGGCTTTCACAGCAGAGACGCGAG	547		
OY	1295	GGGCGGCATCACTCGCGCTGTGGCCCAACGGGCTGGGCAACGCCAAGAACTGAT	1354		
Db	548	GGGCGGCATCACTCGCGCTGTGGCCCAACGGGCTGGGCAACGCCAAGAACTGAT	607		
OY	1355	CACCAAGATGAGCGCGCGGCGGCAAGTACGACTTT	1391		
Db	608	CACCAAGATGAGCGCGCGGCGGCAAGTACGACTTT	644		
RESULT 4					
LOCUS	B1716102	622 bp	mRNA	linear	EST 19-SEP-2001
DEFINITION	1031001E05.y1 C. reinhardtii CC-1690, Stress II (normalized),				
ACCESSION	B1716102				
VERSION	B1716102.1				
KEYWORDS	EST.				
SOURCE	Chlamydomonas reinhardtii.				
ORGANISM	Chlamydomonas reinhardtii				
REFERENCE	Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,				
AUTHORS	P., McDemott,J.P., Strager,J., Silflow,C. and Stern,M.				
TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model,				
JOURNAL	Unicellular System for Analyzing Gene Function and Regulation in				
COMMENT	Vascular Plants. Project: 1031				
	Unpublished (2001)				
	Contact: Charles Hauser				
	DCMB Box 91000				

Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers

FEATURES

Source

1. 622
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Wells et al., (2000). Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 102 a 235 c 178 g 107 t
ORIGIN

Query Match

25.6%; Score 614; DB 13; Length 622;

Best Local Similarity

100.0%; Pred. No. 2,4e-114;

Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

17 CAAACACTCTGCGAGGCACTAGCTCAAAACCTTTTCCAAACAGTTTAAACC 76
18 CAAACACTCTGCGAGGCACTAGCTCAAAACCTTTTCCAAACAGTTTAAACC 68
9 CAAACACTCTGCGAGGCACTAGCTCAAAACCTTTTCCAAACAGTTTAAACC 68
77 CCAATTTGGAGGCGGCTCGCAAGCTCGCTCGTCTCATCGACCACTTATTTT 136
69 CCAATTTGGAGGCGGCTCGCAAGCTCGCTCGTCTCATCGACCACTTATTTT 128
137 CTAATATGTAAGAGCGCAAGATGTGCGCGCTGTGTGAAGCTTGGCGCGGTGTC 196
129 CTAATATGTAAGAGCGCAAGATGTGCGCGCTGTGTGAAGCTTGGCGCGGTGTC 188
197 TATTGCGGCGAGCTCTCTGAGGCGGCGGAGAGTGGCCCCCGGCTCGCTCGGAGCGAG 256
189 TATTGCGGCGAGCTCTCTGAGGCGGCGGAGAGTGGCCCCCGGCTCGCTCGGAGCGAG 248
257 CACGCTGCTGTAGAGCTTGGCAACACTTGAAGCGCGCGCGCTAGGCAAGTGGC 316
249 CACGCTGCTGTAGAGCTTGGCAACACTTGAAGCGCGCGCGCTAGGCAAGTGGC 308
317 TTGCGGCGCTGCGGACCCGCTGCGGAGCGCTTTGATGATGTCACAGCGGCTGGC 376
309 TTGCGGCGCTGCGGACCCGCTGCGGAGCGCTTTGATGATGTCACAGCGGCTGGC 368
377 CGAGCTTGGCAAGCCCAAGAGAGAGCCGAGCGCAAGAGCTGTCGAGAGTGGCTCC 436
369 CGAGCTTGGCAAGCCCAAGAGAGAGCCGAGCGCAAGAGCTGTCGAGAGTGGCTCC 428
437 GCGCGTGTGTGCTATTTGCGAGACCTTGGGCTTGGCGCGGCGGCGCACACCCCAA 496
429 GCGCGTGTGTGCTATTTGCGAGACCTTGGGCTTGGCGCGGCGGCGCACACCCCAA 488
497 GAGCTGGCGGAGGCGCTCGCGCGCTTGAAGAGTGTGTTGACAGCTGTTGG 556
489 GAGCTGGCGGAGGCGCTCGCGCGCTTGAAGAGTGTGTTGACAGCTGTTGG 548
557 GCGCGCTGTGTGCTATTTGCGAGACCTTGGGCTTGGCGCGGCGGCGCACACCCCAA 616

DB 549 GCGGCACTGACATCATGAGAGGCGAGCTGCTGCACCGGCTTACCGAGCACT 608
QY 617 GGAGGCCACCGCG 630
DB 609 GGAGGCCACCGCG 622

RESULT 5

B1999849 617 bp mRNA linear EST 25-OCT-2001

LOCUS

1031079E04.x2 C. reinhardtii CC-1690, Stress II (normalized),

DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

B1999849.1 GI:16434623

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031

JOURNAL

COMMENT

Unpublished (2001)

Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. 617

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"

/note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Wells et al., (2000). Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 114 a 205 c 204 g 93 t 1 others

ORIGIN

Query Match

25.5%; Score 612.8; DB 13; Length 617;

Best Local Similarity

99.5%; Pred. No. 4.3e-114;

Matches 614; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

608 CGAGCACTGAGAGGCGGACCCGCACTCCGAGCGGCGGCTGCGCATGTTACACAGTCTG 667
1 CGAGCACTGAGAGGCGGACCCGCACTCCGAGCGGCGGCTGCGCATGTTACACAGTCTG 60
QY 668 CCGCGCTGTGTGCTATTTGCGAGACCTTGGGCTTGGCGCGGCGGCGCACACCCCAA 727
DB 61 CCGCGCTGTGTGCTATTTGCGAGACCTTGGGCTTGGCGCGGCGGCGCACACCCCAA 120
QY 728 CTGCAAGAGCCCGAGATGATGCTGCGGCGCATGCTCAAGTCTTACCTAGCGGAGGAA 787

Db 121 CTGCAAGAGCCCCCAGATGCTGGCGGCATGTCMACTCTACCTAGCGGAAAAA 180
 Qy 788 GGGGATCGGCGCCAAAGAGATGTCATGTCATCATGCTCCGACGCGCAAGCAGTC 847
 Db 181 GGGGATCGGCGCCAAAGAGATGTCATGTCATCATGCTCCGACGCGCAAGCAGTC 240
 Qy 848 GGAAGCTGACCGGACGTCGTTCTGTGTGAAGCGCGCAACCCCTGCGCCAGCTGACCA 907
 Db 241 GGAAGCTGACCGGACGTCGTTCTGTGTGAAGCGCGCAACCCCTGCGCCAGCTGACCA 300
 Qy 908 CGTCATCACCAAGCTGAGCTGGGCAACATCTCAAGAGAGCGGCGCATCAACTGGCCCA 967
 Db 301 CGTCATCACCAAGCTGAGCTGGGCAACATCTCAAGAGAGCGGCGCATCAACTGGCCCA 360
 Qy 968 GCTGCGCGAGGAGGAGTGGGACATCCAAATGGCGCTGGGCGCGCGCTGCTGTT 1027
 Db 361 GCTGCGCGAGGAGGAGTGGGACATCCAAATGGCGCTGGGCGCGCGCTGCTGTT 420
 Qy 1028 CGGACACACCGGCGGTCATGAGAGCGGCGCTGGCGACGCGCTATGAGCTTTCAAGG 1087
 Db 421 CGGACACACCGGCGGTCATGAGAGCGGCGCTGGCGACGCGCTATGAGCTTTCAAGG 480
 Qy 1088 CAGCGCGCTGCGCGCTGAGCTGAGCGAGGTCGCGGCGGATGAGAGCGCATCAAGAGAC 1147
 Db 481 CAGCGCGCTGCGCGCTGAGCTGAGCGAGGTCGCGGCGGATGAGAGCGCATCAAGAGAC 540
 Qy 1148 CAACATCACCATGTCGCGCGCGCGGCTCAAGTTTGAAGAGCTGTCGAAGACCGCGC 1207
 Db 541 CAACATCACCATGTCGCGCGCGCGGCTCAAGTTTGAAGAGCTGTCGAAGACCGCGC 600
 Qy 1208 CGCGCGCGCGCGCGAGG 1224
 Db 601 CGTCGTCGCGCGCGAGG 617

RESULT 6
 BE724427 601 bp mRNA linear EST 14-SEP-2000
 LOCUS 894076B10.Y1 C. reinhardtii CC-1690, normalized, lambda zap II
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BE724427
 VERSION BE724427.1 GI:10125723
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Chlamydomonas reinhardtii.
 Buiaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales;
 Chlamydomonadaceae: Chlamydomonas.
 1 (bases 1 to 601)
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Jeebvre, P.,
 McDermott, J. P., Silflow, C., Stern, D., and Surzycki, R.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants, Project Phase 2
 Unpublished (2000)
 CONTACT: Charles Hauser
 DCMB Box 91000
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.
 FEATURES
 source
 1. 601
 Location/Qualifiers
 /organism="Chlamydomonas reinhardtii;"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, lambda zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in

ambient levels of CO2 and HS medium bubbled with 5% CO2.
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with Exsist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldi et al (1996) Genome Research 6: 791-806."

BASE COUNT 98 a 225 c 169 g 107 t 2 others
 ORIGIN
 Query Match 24.9%; Score 596.4; DB 12; Length 601;
 Best Local Similarity 99.5%; Pred. No. 9e-111;
 Matches 597; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 17 CAAACACTCTCGAGGACTAGCTCAACCCCTCGAACCTTTTCCACAGTTACAC 76
 Db 2 CAAACACTCTCGAGGACTAGCTCAACCCCTCGAACCTTTTCCACAGTTACAC 61
 Qy 77 CCAATTGAGACGCGCTCCAGCTGCTCGTTGCTCTTCATTCGACCACTATTATT 136
 Db 62 CCAATTGAGACGCGCTCCAGCTGCTCGTTGCTCTTCATTCGACCACTATTATT 121
 Qy 137 CTATATCTGAGACGCGACAGATGTCGCGCTCTGCTGAAGCCCTGCGCGCTGTC 196
 Db 122 CTATATCTGAGACGCGACAGATGTCGCGCTCTGCTGAAGCCCTGCGCGCTGTC 181
 Qy 197 TATTGCGGCGACTCTCTGAGGCGCGGAGGTCGCGCGCTGCTGCGGCGCTGTC 256
 Db 182 TATTGCGGCGACTCTCTGAGGCGCGGAGGTCGCGCGCTGCTGCGGCGCTGTC 241
 Qy 257 CACCGTGCCTGAGACCTTCAACATTGAGGCGCGCGACGCGCTGAGCAAGCTGCG 316
 Db 242 CACCGTGCCTGAGACCTTCAACATTGAGGCGCGCGACGCGCTGAGCAAGCTGCG 301
 Qy 317 TTGCGGCGCTGCGGCAACCGCTGCGGAGCGCTTTGAGTCATGTCAGAGCGCTGCG 376
 Db 302 TTGCGGCGCTGCGGCAACCGCTGCGGAGCGCTTTGAGTCATGTCAGAGCGCTGCG 361
 Qy 377 CGAGCTTGGCAAGCCCAAGAGAGAGCCCAAGGCGGCGGCGGCTGAGTGGCTGCG 436
 Db 362 CGAGCTTGGCAAGCCCAAGAGAGAGCCCAAGGCGGCGGCGGCTGAGTGGCTGCG 421
 Qy 437 GGCCTGCTGTCGCTATTGCGGAGACCTGCGGCTGCGGCGGCGGCGGCGGCGGCA 496
 Db 422 GGCCTGCTGTCGCTATTGCGGAGACCTGCGGCTGCGGCGGCGGCGGCGGCGGCA 481
 Qy 497 GCAGCTGCGGAGGCGCTGCGGCGCTGCGGCTTTCAGAGGTTTGAACGCTGTTGG 556
 Db 482 GCAGCTGCGGAGGCGCTGCGGCGCTTTCAGAGGTTTGAACGCTGTTGG 541
 Qy 557 CGCGGACCTGACCATATGAGAGGAGGCGGAGGCTGCTGACCGGCTCAGCAGACCT 616
 Db 542 CGCGGACCTGACCATATGAGAGGAGGCGGAGGCTGCTGACCGGCTCAGCAGACCT 601

RESULT 7
 AV389547 558 bp mRNA linear EST 29-SEP-2000
 LOCUS AV389547 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
 DEFINITION cDNA clone CM044d10_r, mRNA sequence.
 ACCESSION AV389547
 VERSION AV389547.1 GI:6543763
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Chlamydomonas reinhardtii.
 Buiaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales;
 Chlamydomonadaceae: Chlamydomonas.
 1 (bases 1 to 558)
 Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabara, S.
 A large scale structural analysis of cDNAs in a unicellular green
 alga, Chlamydomonas reinhardtii. I. Generation of 3433
 non-redundant expressed sequence tags

JOURNAL MEDLINE	DNA Res. 6 (6), 369-373 (1999)
COMMENT	Contact: Yasukazu Nakamura The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers 1..558 /organism="Chlamydomonas reinhardtii" /strain="C9" /db_xref="taxon:3055" /clone_1ib="CMO44d10_r" /clone_1lb="Chlamydomonas reinhardtii C9" /dev_stage="photoautotrophic growth" /note="Vector: pBluescriptII SK; Site_1: EcoRI; Site_2: XhoI"
FEATURES	
SOURCE	
BASE COUNT	113 A 171 C 202 G 72 T
ORIGIN	
Query Match	23.2%; Score 556.4; DB 10; Length 558;
Best Local Similarity	99.8%; Pred. No. 1.1e-102;
Matches 557; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	1213 CGCGGCGCGAGGCCGGCCGGCACACCCCCGGGCGCTGGCTTGGGACGGCGCGCG 1272
Db	1 CGCGGCGCGAGGCCGGCCGGCACACCCCCGGGCGCTGGGACGGCGCGCGCG 60
OY	1273 GCTTCACCAGCAGAGACGGCGAGCGGCATCATCACTCGGTGGCCGTGGCCAAAGGCG 1312
Db	61 GCTTCACCAGCAGAGACGGCGAGCGGCATCATCACTCGGTGGCCGTGGCCAAAGGCG 120
OY	1333 TGGGCAAGCCCAAAGAAGCTGATCATCAAGATCAGAGCCGGGAGGCAAGTAGACACTTTG 1392
Db	121 TGGGCAAGCCCAAAGAAGCTGATCATCAAGATCAGAGCCGGGAGGCAAGTAGACACTTTG 180
OY	1393 TGGAATCATGGCTTGCCCGCGGGGCTGTGTGGGGGGCGGCGGACGCCGCTCCACCG 1452
Db	181 TGGAATCATGGCTTGCCCGCGGGGCTGTGTGGGGGGCGGCGGACGCCGCTCCACCG 240
OY	1453 ACAAGGCGCATCAACGAGAGCGGACGAGCGCGCTGTACAACCTGGAGCAGAAAGTCCACGC 1512
Db	241 ACAAGGCGCATCAACGAGAGCGGACGAGCGCGCTGTACAACCTGGAGCAGAAAGTCCACGC 300
OY	1513 TGCGCGCGAGCCACGAGAACCCGTCATCCGGAGCTGTACGACACGTACCTCGGAGAGC 1572
Db	301 TGCGCGCGAGCCACGAGAACCCGTCATCCGGAGCTGTACGACACGTACCTCGGAGAGC 360
OY	1573 CGCTGGGCGCAAGAGCGCAGAGCTGCTCACACCCACTACCTGGCCGGCGCGCTGGAG 1632
Db	361 CGCTGGGCGCAAGAGCGCAGAGCTGCTCACACCCACTACCTGGCCGGCGCGCTGGAG 420
OY	1633 AGAAGACGAGAAAGATGAGAGCGCCGAGAGCTTTTGGCGGAGACAGTTCGAAGC 1692
Db	421 AGAAGACGAGAAAGATGAGAGCGCCGAGAGCTTTTGGCGGAGAGACAGTTCGAAGC 480
OY	1693 GAAGGGGCGATTAGAGAGTACCGTAATAATATGCACTAGTGGTGAATCGGGGTCTCTCTT 1752
Db	481 GAAGGGGCGATTAGAGAGTACCGTAATAATATGCACTAGTGGTGAATCGGGGTCTCTCTT 540
OY	1753 TATATTGAATGGGTCMA 1770
Db	541 TATATTGAATGGGTCMA 558
RESULT 8	
BG848385	602 bp mRNA linear EST 29-MAY-2001
LOCUS	1024021B05.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION	Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION	BG848385
VERSION	BG848385.1 GI:14229569
KEYWORDS	EST.

SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.
REFERENCE	1 (bases 1 to 602) Grossman, A., Davies, J., Federlepiel, N., Harris, E., Lefebvre, P., Mcdermott, J. P., Silflow, C., Stern, D. and Surzycki, R. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2 Unpublished (2000)
JOURNAL	Contact: Charles Hauser
COMMENT	DCMB Box 91000 Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chausere@duke.edu. Location/Qualifiers
FEATURES	source 1..602 /organism="Chlamydomonas reinhardtii" /strain="CC-1690 wild type mt+ 21gr" /db_xref="taxon:3055" /clone_1lb="C. reinhardtii CC-1690, normalized, Lambda Zap II" note="Vector: pBluescript II SK-; Site, 1: EcoRI; Site, 2: XhoI; This library, constructed by John Davies and Jeffrey Mcdermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phase. The library was normalized using method 4 described in Donato et al (1996) Genome Research 6: 791-806."
BASE COUNT	134 a 141 c 184 g 143 t
ORIGIN	
Query Match	22.7%; Score 545; DB 12; Length 602;
Best Local Similarity	95.7%; Pred. No. 2.4e-100;
Matches 560; Conservative	0; Mismatches 25; Indels 0; Gaps 0;
Dn	1800 TCAGTCGTCTCACACATGGGGCACTGCGTGCAGGCCAGTAGGCTGTTCACACAGCT 1859
Qy	18 TCCTTGCTCTTTCCACATTGTTGTGACGGGGCTCCGATAGGCTGTTCACACAGCT 77
Dn	1860 GGCAATTAGCGGTAGTA CTGGCATGAGGGAAGCGGGCTTGCTAA CGAATGGCGTATCC 1919
Qy	78 GGCATTAGCGGTAGTACTGGCATGAGGAGCGGGGCTTGCTAACCGAATGGCGTATCC 137
Dn	1920 TCCAGGGCACTGGGAATGGCGCGTGCCTCATACAGCAAATTTCTTGGCTTCATCGCTTC 1979
Qy	138 TCCAGGGCACTGGGAATGGCGCGTGCCTCATACAGCAAATTTCTTGGCTTCATCGCTTC 197
Dn	1980 TGGATTTGAAGCTCACAAAACCTGCATTCTAATTGCTGTGTTTACAGTGCCTCAATCTT 2039
Qy	198 TGGATTTGAAGCTCACAAAACCTGCATTCTAATTGCTGTGTTTACAGTGCCTCAATCTT 257
Dn	2040 GGTTGAAGCTAAACATGTTTGGGAACAATTCATCTTAAGAAGCGTGTGGGGGTTGAGG 2099
Qy	258 GGTTGAAGCTAAACATGTTTGGGAACAATTCATCTTAAGAAGCGTGTGGGGGTTGAGG 317
Dn	2100 ATGCGCAGCTTGCGCTGCTGGTGGGGGGAACGTGGTNAAGCTTTAGGCTAGCTGCGC 2155
Qy	318 ATGCGCAGCTTGCGCTGCTGGTGGGGGGAACGTGGTNAAGCTTTAGGCTAGCTGCGC 377
Dn	2160 ATTAGCAACGGGGCCCCGGAGGATTGAGCACTTGACTCGGAACCTTAATGAACGTAGGCG 2215
Qy	378 ATTAGCAACGGGGCCCCGGAGGATTGAGCACTTGACTCGGAACCTTAATGAACGTAGGCG 437

QY 2220 TTATATCCACCGTATGCGCATTTGATGAGCAACCGCGGTAGGAAGCGGAG 2279
 DB 438 TTATATCCACCGTATGCGCATTTGATGAGCAACCGCGGTAGGAAGCGGAG 497
 QY 2280 AGATGATTCGCAACCGCTTAAAGAACGGCATAGTACTATGACATCTTGATGAGC 2339
 DB 498 AGATGATTCGCAACCGCTTAAAGAACGGCATAGTACTATGACATCTTGATGAGC 557
 QY 2340 CTGGGCGACGCGACGAGAGAGGTGTCATCAGCGCTTGTA 2384
 DB 558 CTGGGCGACGCGACGAGAGAGGTGTCATCAGCGCTTGTA 602

RESULT 9 599 bp mRNA linear EST 19-SEP-2001
 LOCUS B1716925 1031013611.y1 C. reinhardtii CC-1690, Stress II (normalized),
 DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION B1716925 GI:15692620
 VERSION B1716925.1
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE 1 (bases 1 to 599)
 Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
 P., McDermott, J.P., Shrager, J., Silflow, C., and Stern, D.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 1031

TITLE

JOURNAL Contact: Charles Hauser
 DCMB Box 91000
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.

COMMENT

FEATURES
 source location/Qualifiers

1. 599
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress II (normalized)
), Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; Stress condition II library, constructed by John
 Davies and Jeffrey McDermott, combines cDNAs from CC-1690
 cells grown to mid-log phase in TAP (NH4+ - containing)
 and shifted to TAP - NO3- (24hrs); H2 production
 conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
 Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
 sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
 sites. pBluescript II SK- plasmids were excised from the
 lambda Zap clones by superinfection with ExAssist
 (Stratagene) phage. The library was normalized using
 method 4 described in Bonaldi et al., (1996) Genome
 Research 6: 791-806."
 Research 6: 791-806." 97 t

BASE COUNT 123 a 172 c 207 g 97 t

Query Match 22.7%; Score 543.8; DB 13; Length 599;
 Best Local Similarity 99.6%; Pired. No. 4.1e-100;
 Matches 545; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1298 CGGCATCACACTGGCGGCGCGGCGCAACCGGCTGCGCAACCGCAAGACGTGATC 1357
 DB 1 CGGCATCACACTGGCGGCGGCGCGGCGCAACCGGCTGCGCAACCGCAAGACGTGATC 60

QY 1358 CAAGATGACAGCGCGGCGAGCGCAAGTACGACTTTGTGAGATATGCGCTCGCGGG 1417
 DB 61 CAAGATGACAGCGCGGCGAGCGCAAGTACGACTTTGTGAGATATGCGCTCGCGGG 120
 QY 1418 CTGTGTGGGCGGGGCGGCGCGCGCTTCACCGCAAGGCCATACCGAAGAGCGGCA 1477
 DB 121 CTGTGTGGGCGGGGCGGCGCGCGCTTCACCGCAAGGCCATACCGAAGAGCGGCA 180
 QY 1478 GGGCGCGCTGTATCAACCTGGAAGAGATGACGCTGCGCGCGGCGGCGAGAACCGGTC 1537
 DB 181 GGGCGCGCTGTATCAACCTGGAAGAGATGACGCTGCGCGCGGCGGCGAGAACCGGTC 240
 QY 1538 CATCCGCGAGCTGTACGACAGTACTCGGAGAGCGGCTGCGGCGACAGGCGGCGAGCT 1597
 DB 241 CATCCGCGAGCTGTACGACAGTACTCGGAGAGCGGCTGCGGCGACAGGCGGCGAGCT 300
 QY 1598 GCTGCACACCCACTTACGTGCGCGCGGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAG 1657
 DB 301 GCTGCACACCCACTTACGTGCGCGCGGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 1658 GCCAGAGGCTCTTTGGCGCGGAGACAGCTTCAAGCGAGGCGGCGGCTATTAGCAGTACCGTA 1717
 DB 361 GCCAGAGGCTCTTTGGCGCGGAGACAGCTTCAAGCGAGGCGGCGGCTATTAGCAGTACCGTA 420
 QY 1718 AATATGACATGATGCGGTGATGCGGCTGCTCCTTTATATTGAATGGGCTCAAAATAGGC 1777
 DB 421 AATATGACATGATGCGGTGATGCGGCTGCTCCTTTATATTGAATGGGCTCAAAATAGGC 480
 QY 1778 GGGGCGTCAAAATGTTTCTTTTGAATGCTGTACAGCATGGGCGACGTTGCGAGGCC 1837
 DB 481 GGGGCGTCAAAATGTTTCTTTTGAATGCTGTACAGCATGGGCGACGTTGCGAGGCC 540
 QY 1838 AGTAGGC 1844
 DB 541 AGTTGCC 547

RESULT 10 621 bp mRNA linear EST 29-MAY-2001
 LOCUS BG848384/c 1024021E05.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BG848384 GI:14229568
 VERSION BG848384.1
 KEYWORDS EST.

SOURCE

ORGANISM Chlamydomonas reinhardtii.
 Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE

1 (bases 1 to 621)
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
 McDermott, J.P., Silflow, C., Stern, D., and Surzycki, R.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2

JOURNAL

COMMENT Unpublished (2000)
 Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.

FEATURES

source location/Qualifiers

1. 621
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
 II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells, grown to

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 499)
Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S.
A large scale structural analysis of cDNAs in a unicellular green
alga, Chlamydomonas reinhardtii. I. Generation of 3433
non-redundant expressed sequence tags

JOURNAL
MEDLINE
COMMENT
20152988
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES

source

1..499
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CL58a12_r"
/clone_1ib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 77 a 191 c 141 g 90 t
ORIGIN

Query Match 20.8%; Score 499; DB 10; Length 499;
Best Local Similarity 100.0%; Pred. No. 4.8e-91;

Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

40 CTCGAACCTCGAATCTTTTCCACAGTTTACACCCCAATTCGAGCCGCTCCACG 99
1 CTCGAACCTCGAATCTTTTCCACAGTTTACACCCCAATTCGAGCCGCTCCACG 60
100 TCGCTCGGTCCTCTTCAATCGCACCACTATTATTCTTAATTCGTAGACCGACAGA 159
61 TCGCTCGGTCCTCTTCAATCGCACCACTATTATTCTTAATTCGTAGACCGACAGA 120
160 TGTGGGCGCTCGTGAAGCCCTGCGGCGGCTGTCTATTTCGCGGACGCTCTGAGGG 219
121 TGTGGGCGCTCGTGAAGCCCTGCGGCGGCTGTCTATTTCGCGGACGCTCTGAGGG 180
220 CCGGAGGAGTCGCGCCCGCGGCTCCGCTCGACGACGACCGGTGATAGCCCTTGCA 279
181 CCGGAGGAGTCGCGCCCGCGGCTCCGCTCGACGACGACCGGTGATAGCCCTTGCA 240
280 CACTTAGAGCGCCGCGACGCGGCTTAGGCAAGCTGCTGCGGAGTGCAGACCGCTG 339
241 CACTTAGAGCGCCGCGACGCGGCTTAGGCAAGCTGCTGCGGAGTGCAGACCGCTG 300
340 CCGAGGCGCTTGAATCATATGTCACAGAGCGCTCCGCGAGCTTGCCAAAGCAAG 399
301 CCGAGGCGCTTGAATCATATGTCACAGAGCGCTCCGCGAGCTTGCCAAAGCAAG 360
400 ACCCGACGCGACGACGTCGTGCTGAGGTGCTCGGCGGCTTGCTGATGCTATTGCG 459
361 ACCCGACGCGACGACGTCGTGCTGAGGTGCTCGGCGGCTTGCTGATGCTATTGCG 420
460 AGACCTTGGGCTGCGGCGCGGCGGCGACCAACCCCAAGACGCTGCGGAGGCTTCGCG 519
421 AGACCTTGGGCTGCGGCGCGGCGGCGGCGACCAACCCCAAGACGCTGCGGAGGCTTCGCG 480
499 GCTCGGCTTTGACGAGGT 538
481 GCTCGGCTTTGACGAGGT 499

RESULT 13
AV386674 498 bp mRNA linear EST 29-SEP-2000
LOCUS AV386674 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
DEFINITION CDNA clone CM004a11_r, mRNA sequence.

ACCESSION AV386674
VERSION AV386674.1 GI:6540890
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

JOURNAL
MEDLINE
COMMENT
20152988
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES

source

1..498
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM004a11_r"
/clone_1ib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 78 a 194 c 139 g 87 t
ORIGIN

Query Match 20.8%; Score 498; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 7.6e-91;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

25 CTCGAGGACATAGCTTCAACCTCGAATCTTTTCCACAGTTTACACCCCAATTCG 84
1 CTCGAGGACATAGCTTCAACCTCGAATCTTTTCCACAGTTTACACCCCAATTCG 60
85 GACGCGCTCCAGCTCGCTCGTGTCTCTTCAATCGCACCACTATTATTCTTAATTC 144
61 GACGCGCTCCAGCTCGCTCGTGTCTCTTCAATCGCACCACTATTATTCTTAATTC 120
145 GTAGAGCGGACAGATGTCGGGCTCGTGTGAAGCCCTGCGGCGGCGGCTGATATTCGG 204
121 GTAGAGCGGACAGATGTCGGGCTCGTGTGAAGCCCTGCGGCGGCGGCTGATATTCGG 180
205 GCAGCTCTGACGAGCGCGGACGTCGCTCCCGCGCTTCGCTCGACGACACCGTGC 264
181 GCAGCTCTGACGAGCGCGGACGTCGCTCCCGCGCTTCGCTCGACGACACCGTGC 240
265 GTGTAGCCCTTGCAACATTTGAGGCGCGCGACGCGGCTTAGGCAAGCTGCTTGCGGG 324
241 GTGTAGCCCTTGCAACATTTGAGGCGCGCGACGCGGCTTAGGCAAGCTGCTTGCGGG 300
325 CTGCGGACACCGCTGCGGAGGCGGCTTTGAGTCATGTCAGAGGCGGCTGCGGAGCTTG 384
301 CTGCGGACACCGCTGCGGAGGCGGCTTTGAGTCATGTCAGAGGCGGCTGCGGAGCTTG 360
385 CCAAGCCCAAGAGACACCCACGCGACGTCGTGCTGAGGTGCTTCGCGGCGTTTC 444
361 CCAAGCCCAAGAGACACCCACGCGACGTCGTGCTGAGGTGCTTCGCGGCGTTTC 420
445 GTGTGCTATTGTCGAGACCTTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 504
421 GTGTGCTATTGTCGAGACCTTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
505 CCGAGGCGCTCGCGCGCC 522
481 CCGAGGCGCTCGCGCGCC 498


```
QY 1800 TGAAGTGTGTCAACAGATGGGGACGCTGTGCGGAGGCCAGTAGGCTTTCATCTGCACGCT 1859
Db 522 TGCCTGTCTCTTCCACATTGTGGGCGCGCTCGCATAGGCTGTTCACCTGCACGCT 463
QY 1860 GGCATTAGGCGTATAGTACTGTGCATGAGGGAGCGGGCTTGCTAACCGAATGGCGTATCCC 1919
Db 462 GGCATTAGGCGTATAGTACTGTGCATGAGGGAGCGGGCTTGCTAACCGAATGGCGTATCCC 403
QY 1920 TCCAGGGGCACTGCGAATGGCGCGTGCATCAACGCAAAATCTTGCCCTTCATCGCTTC 1979
Db 402 TCCAGGGGCACTGCGAATGGCGCGTGCATCAACGCAAAATCTTGCCCTTCATCGCTTC 343
QY 1980 TGGATATTGAAGCTGCACAAACCTGCATTCTATTGCTTGTTCACAGTGCCTCAATCTT 2039
Db 342 TGGATATTGAAGCTGCACAAACCTGCATTCTATTGCTTGTTCACAGTGCCTCAATCTT 283
QY 2040 GGTGGAAGCTTAACATGTTTGGGAACATTCTTACTTAAGCGTGGGGGTTGAGG 2099
Db 282 GGTGGAAGCTTAACATGTTTGGGAACATTCTTACTTAAGCGTGGGGGTTGAGG 223
QY 2100 ATGCGCAGCTTGTGCGCTGTGGTGGGCGGGAACGTGGGTAGCATTAGGCTAGCTGGC 2159
Db 222 ATGCGCAGCTTGTGCGCTGTGGTGGGCGGGAACGTGGGTAGCATTAGGCTAGCTGGC 163
QY 2160 ATACGACAAAGGGGCGCGTAGGATTGAGCACTTGACTCGCGAACTTATGAACGTAGCGC 2219
Db 162 ATACGACAAAGGGGCGCGTAGGATTGAGCACTTGACTCGCGAACTTATGAACGTAGCGC 103
QY 2220 TTTATCCCAACCGTATGCGATTGACGTTGGTGTAGGCAACGAGCGGTAGGAAGCGGAG 2279
Db 102 TTTATACCCCGTATGCGATTGACGTTGGTGTAGGCAACGAGCGGTAGGAAGCGGGA 43
QY 2280 AGATGATTGCAAAACGCTGTAAAGAAAGACGGCATAG 2315
Db 42 AGATGATTGCAAAACGCTGTAAAGAAAGACGGCATAG 7
```

Search completed: June 8, 2003, 12:43:28
Job time : 3037 secs

This Page Blank (uspto)